

## EAST Search History

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	2442	flavivirus	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:33
L2	7135	envelope adj protein	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:33
L3	681	I1 and I2	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:33
L4	1473	domain adj III	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:34
L5	37	I3 and I4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:34
L6	1065793	inhibit\$	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:34
L7	31	I5 and I6	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:34
L8	246025	antibod\$	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:34
L9	31	I7 and I8	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:35

## EAST Search History

L10	168742	ligand	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:35
L11	17	I9 and I10	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/07/09 20:35



A service of the National Library of Medicine  
and the National Institutes of Health

[My NCBI](#) [\[Sign In\]](#) [\[Register\]](#)

All Databases    PubMed    Nucleotide    Protein    Genome    Structure    OMIM    PMC    Journals    Books

Search

for

Limits    Preview/Index    **History**    Clipboard    Details

- Search History will be lost after eight hours of inactivity.
- To combine searches use # before search number, e.g., #2 AND #6.
- Search numbers may not be continuous; all searches are represented.
- Click on query # to add to strategy

[About Entrez](#)

[Text Version](#)

[Entrez PubMed](#)

[Overview](#)

[Help | FAQ](#)

[Tutorials](#)

[New/Noteworthy](#)

[E-Utilities](#)

Search	Most Recent Queries	Time	Result
	#24 Search flavivirus domain III	20:39:18	<a href="#">58</a>
	#23 Search flavivirus	20:39:15	<a href="#">8529</a>

[PubMed Services](#)

[Journals Database](#)

[MeSH Database](#)

[Single Citation Matcher](#)

[Batch Citation Matcher](#)

[Clinical Queries](#)

[Special Queries](#)

[LinkOut](#)

[My NCBI](#)

[Related Resources](#)

[Order Documents](#)

[NLM Mobile](#)

[NLM Catalog](#)

[NLM Gateway](#)

[TOXNET](#)

[Consumer Health](#)

[Clinical Alerts](#)

[ClinicalTrials.gov](#)

[PubMed Central](#)

Jul 6 2006 07:09:19

TJL

FAPELANNTFVIDGPETEECPTANRAWNSMEVEDFGFGLTSTRMFLRIRETNTECDSD  
KIIGTAVKNMAVHSDSLWSYWIESTGLNDTWKLERAVLGEVKSCWTPEHTLWGDBGLES  
DLIIPITLAGPRSNHNRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTISDSCGHRGPA  
ARTTTESGKLITDWCCRSCTLPPLRFQTENGWYGMIEPRTDHDEKTLVQSRVNAYNA  
DMIDPFQLGLLVVFLATQEVLRKRWTAKISIPAIMLALLVLVFGGITYTDVLRYVILV  
GAAFAEANSGGDVVHLALMATFKIQPVFLVASFLKARWTNQESILLMLAAFFQMAYY  
DAKNVLSWEVDVLNSLVAWMILRAISFTNTSNVVVPLALLTPGLKCLNLDVYRIL  
LLMVGVS LIKEKRSSAAKKGACLICLALASTGVFNPMILAAGLMACDPNRKRGWPA  
TEVMTAVGLMFAIVGGLAELDIDSMAPMTIAGLMFAFVISGKSTDMWERTADITW  
ESDAEITGSSEVDVRLDDGNFOLMNDPGAPWKIWMRMACLAISAYTPWAILPSVI  
GFWITLQYTAKRGGVWLDTSPSPKEYKKGDTTGVYRIMTRGLLGSYQAGAGVMVEGVFH  
TLWHTTKGAALMSGEGRLDPYWSVKEDRLCYGGPKLQHKWNHGDEVQMIIVEPGKN  
VKNVQTKPGVFKTPEGEIGAVTLDYPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS  
AIVQGERMEEPAPAGEPEMLRKQITVLDLHPGAGKTRKILPQIKEAINKRLRTAV  
LAPTRVVAEAMSEALRGLPIRYQTSAVHREHSGNEIVDVMCHATLTHRLMSPHRVPNY  
NLFIIMDEAHFTDPASIAARGYIATKVELGEAAAI FMTATPPGTSDFPESNAPITDMQ  
TEIPDRAWNTGYEWITEYVGKTVWFVPSVKMGNEIALCLQRAGKKVIQLNRKSYETEY  
PKCKNDWDVFITTDISEMGANFKASRVIDSRKSVKPTII EEGDGRVILGEPSAITAA  
SAAQRRGRIGRNPSQVGDECYGHTNEDSNFAHWTTEARIMLDNINMPNGLVAQLYQ  
PEREKVYTMGEYRLGEERKNFLEFLRTADLPVWLAYKAAAGISYHDRKWCDFGPR  
TNTILEDNNNEVEVITKLGERKILRPRWADARVYSDHQALKSFKDFASGKRSQIGLVEV  
LGRMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALELPDALQTIALLSVMSLG  
VFVLLMQRKGIGKIGLGGVILGAATFFCWMADVPGTKIAGMLLSLLMIVLYPEPEK  
QRSQTDNQLAVFLICVLTLSAVAANEMGWLDKTKNDISSLGHKPEARETTLGVESF  
LLDLRPATAWSLYAVTAVLTPLKHLITSDYINTSLTSINVQASALFTLARGFPFVD  
VGVSALLAAGCWGQVTLTVTAAALLFCHYAYMVGWQAEAMRSAQRRTAACIMKN  
AVVDGIVATDVPELERTTPVMQKVGQIMLI LVSMAAVVNPNSVRTVREAGILTTAAA  
VTLWENGASVWNATTAAIGLCHIMRGGWLSCLSITWTLIKNMEEKPLKRGGAKGRTLG  
EVWKERLNHMTKEEFTTRYRKEAITEVDRSAKHARREGNITGGHPVSRGTAKLRWLVE  
RRFLEPVGVKVDLGCGRGGWCYYMATQKRVQEVKGTYKGGPGHEEPQLVQSYGWNIVT  
MKSGVDVFYRPSEASDTLLCDIGESSSSAEVEEHTVRVLEMVEDWLHRGPKEFCIKV  
LCPYMPKVIEMETLQRRYGGGLVRNPLSRNSTHEMYWVSHASGNIVHSVNMITSQVLL  
GRMEKKTWKGPQFEEDVNLGSGTRAVGKPLNSDTSKIKNRIERLKKEYSSTWHQDAN  
HPYRTWNYHGSYEVKPTGSASSLNVVRLSKPWDTITNVTTMAMTDTPFGQQRVF  
KEKVDTKAPEPPEGVKYVVLNETTNWLWAFLARDKKPRMCREEFIGKVNNSAALGAMF  
EEQNQWKNAREAVEPDPKFWEMVDEERE AHLRGECNTCIYNMMGKREKKPGEFGKAKGS  
RAIWFMWL GARFLEFEALGFLNEDHWLGRKNSGGVEGLQKLYILKEVGTKPGGK  
IYADDTAGWDTRITKADLENEAKVLELLDGEHRRLARSIIELTYRHKKVVKVMRPAADG  
KTVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLGKGKPK  
VRTWLFENGEERLSRMMAVSGDDCVVKPLDRFATSLHFLNAMS KVRKDQEWKPSTGW  
YDWQQVPFCNSNHFTELIMKDGR TLVPCRGQDELIGRARISPAGWNVRDTACLAKSY  
AQMWLLYFRRDLRLMANAICS AWPVNWVPTGRTTWSIHAKGEWTTEDMLS VWNRV  
WIEENEWMEDKTPVERSDVPYSKGREDIWCGLIGTRTRATWAENIHVA INQVRSVI  
GEEKYVDYMSLRLRYEDTIVVEDTIVL"

mat\_peptide  
97. .411  
/product="C protein"  
mat\_peptide  
412. .966  
/product="prM protein"  
sig\_peptide  
412. .465  
/note="prM signal peptide"  
mat\_peptide  
466. .741  
/product="cleaved amino terminal prM fragment"  
mat\_peptide  
742. .966  
/product="M protein"  
sig\_peptide  
964. .2472  
/note="E signal peptide"  
mat\_peptide  
967. .2469

```

    /product="E protein"
mat_peptide      2470. .3525
    /product="NS1 protein"
mat_peptide      3526. .4218
    /product="NS2A protein"
mat_peptide      4219. .4611
    /product="NS2B protein"
mat_peptide      4612. .6468
    /product="NS3 protein"
mat_peptide      6469. .6915
    /product="NS4A protein"
sig_peptide      6847. .6915
    /note="2K peptide"
mat_peptide      6916. .7683
    /product="NS4B protein"
mat_peptide      7684. .10398
    /product="NS5 protein"
misc_feature     2398. .2469
    /note="NS1 signal peptide"
3'UTR           10402. .11048

```

ORIGIN

```

Query Match          87.4%;  Score 1327.8;  DB 10;  Length 11048;
Best Local Similarity 92.3%;  Pred. No. 0;
Matches 1398;  Conservative  0;  Mismatches 117;  Indels  0;  Gaps   0;

Qy      1 CGGAATTCACTGTTAGGAATGAGCAACAGGGACTTCTGGAGGGAGTGTCTG 60
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      956 CAGCATACAGCTTAACTAGGAAATGAGCAACAGAGACTTCTGGAGGGAGTGTCTG 1015
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      61 GAGCTACATGGGTTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAAG 120
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1016 GAGCCACATGGGTTGATCTGGTACTGGAAGGCGACAGCTGTGTAACCATAATGTCAAAAG 1075
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      121 ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAACATCGCAGATGTGC. 180
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1076 ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAACATCTGCAGATGTGC 1135
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      181 GTAGCTACTGCTACTTAGCTTCGGTCAGTGATCTGTCAACAAAAGCCCGTGTCCAACCA 240
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1136 GCAGTTACTGTTACCTAGCTTCAGTCAGTGACTTGTCACAAAGAGGCCGTGTCCAACCA 1195
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      241 TGGGTGAAGCTCACACGAGAAAAGAGCCGACCCCTGCCTTGTTGCAAGCAAGGCGTCG 300
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1196 TGGGTGAAGCCCACAATGAAAAAGAGCTGATCCCGCCTCGTTGCAAGCAAGGCGTTG 1255
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      301 TAGACAGAGGATGGGGAAATGGATGCGGACTGTTGAAAGGGGAGCATTGACACATGTG 360
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1256 TAGATAGAGGATGGGGAAACGGATGCGGACTGTTGAAAGGGAGCATTGACACATGTG 1315
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      361 CAAAGTTGCCGTACAACCAAGGCACTGGTTGGATTATCCAGAAGGAAAACATCAAGT 420
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1316 CGAAGTTGCCGTACAACCAAAGCGACTGGTTGGATCATCCAGAAGGAAAACATCAAGT 1375
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      421 ACGAGGTTGCCATATTGTGCATGGCCCGACGACTGTCGAATCACATGGCAATTATTCAA 480
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1376 ATGAGGTTGCCATATTGTGCATGGCCCGACGACCCTGAATCTCATGGTGATTATTCAA 1435

```

Qy	481	CACAGATAGGGCTACCAAGCAGGAAGGTCAGCATAACTCCATCGGCACCACCTACA	540
Db	1436	CACAGATAGGGCCACCCAGGCTGGAAGATTAGCATAACTCCATCGGCACCATCTTACA	1495
Qy	541	CGCTGAAGTTGGGTGAGTATGGTGAGGTACAGTTGACTGTGAGGCCACGGTCAGGAATAG	600
Db	1496	CGCTAAAGTTGGGTGAGTATGGTGAGGTAAACGGTGATTGTGAGGCCACGGTCAGGAATAG	1555
Qy	601	ACACTAGCGCTTACTACGTTATGTCAGTGGGTGCGAAGTCCTCTGGTCACCGAGAAT	660
Db	1556	ACACTAGCGCTTACTACGTTATGTCAGTGGTGCGAAGTCCTCTGGTCACCGAGAAT	1615
Qy	661	GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG	720
Db	1616	GGTTCATGGATCTGAACCTGCCATGGAGCAGTGCTGGAAGCACCACGTGGAGGAATCGGG	1675
Qy	721	AAACACTGATGGAGTTGAAGAACCTCATGCCACCAAACAATCTGTCGTAGCTCTAGGGT	780
Db	1676	AAACACTGATGGAGTTGAAGAACCTCATGCCACCAAACGATCTGTTGTGGCTCTAGGGT	1735
Qy	781	CGCAGGAAGGTGCCTTGCACCAAGCTCTGGCTGGAGCAATTCTGTTGAGTTCTCAAGCA	840
Db	1736	CGCAGGAAGGCCTTGCACCAAGCTCTGGCGGAGCGATTCTGTTGAATTCTCAAGCA	1795
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTAAGATGGAGAAGTTGCAGC	900
Db	1796	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTAAGATGGAGAAGTTGCAGC	1855
Qy	901	TGAAGGAAACAACATATGGTGTATGCTCAAAAGCATTCAAATTGCTAGGACTCCGCTG	960
Db	1856	TGAAGGAAACAACATACGGAGTATGTTCAAAAGCGTTCAAATTGCTGGACTCCTGCTG	1915
Qy	961	ACACTGGTCATGGAACGGTGGTGTGGAACTGCAGTATACCGAAAAGACGGCCTTGCA	1020
Db	1916	ACACTGGCATGGAACGGTGGTGTGGAACTGCAGTACACCGAACGGACGGTCCCTGCA	1975
Qy	1021	AAGTGCCATTCTCTGTGGCTCCCTGAACGACCTACACCGTTGGAAGGCTGGTGA	1080
Db	1976	AAGTGCCATTCTCTCGTAGCTCCCTGAATGACCTCACACCTGTTGGAAGACTGGTAA	2035
Qy	1081	CTGTGAATCCATTGTGTCTGTGGCTACGCCACTCGAAGGTTTGATTGAACCGAAC	1140
Db	2036	CAGTGAATCCATTGTGTCTGTGGCCACGCCACTCGAAGGTTTGATTGAACCGAAC	2095
Qy	1141	CCCCGTTAGTGAECTTACATCGTGGTGGGGAGAGGAGAACAGCAGATAAACCAACT	1200
Db	2096	CCCCGTTGGTGAECTCTACATCGTGGTGGGAAGAGGAGAACAGCAGATAAACCAACT	2155
Qy	1201	GGCACAAATCTGGAGCAGTATTGAAAGGCTTCACCACTACACTCAGAGGAGCTAAC	1260
Db	2156	GGCACAAATCCGGGAGCAGCATTGAAAGGCTTACTACACACTCAGAGGAGCTAAC	2215
Qy	1261	GACTTGCAGCTCTGGAGACACTGCCTGGGATTTGGATCAGTCGGAGGGTTTCACCT	1320
Db	2216	GACTTGCAGCTCTGGAGACACTGCTGGGATTTGGATCAGTCGGAGGGTATTACCT	2275

Qy	1321 CGGTAGGGAAAGCCATACACCAAGTTTGGAGGAGCCTTAGATCACTCTTGAGGGA 1380 
Db	2276 CGGTGGGGAAAGCTATACACCAAGTCTTGAGGAGCTTAGATCACTTTGGAGGGA 2335
Qy	1381 TGCCTGGATCACACAGGGCTCTGGAGCTCTGCTGTGGATGGAATTACGCC 1440 
Db	2336 TGCCTGGATCACACAGGGACTCTGGAGCTCTGTTGTGGATGGAATCAATGCC 2395
Qy	1441 GTGACAGGTCAATTGCTATGACGTTCCCTGCGGTTGGAGGAGTCTGCTCTCCTTCGG 1500 
Db	2396 GTGACAGGTCAATTGCTATGACGTTCCCTGCGGTTGGAGGAGTTGCTCTCCTTCGG 2455
Qy	1501 TCAACGTCCATGCTG 1515 
Db	2456 TCAACGTCCACGCTG 2470

#### RESULT 4

AY532665

LOCUS AY532665 11038 bp RNA linear VRL 09-DEC-2004  
 DEFINITION West Nile virus strain B956 polyprotein gene, complete genome.  
 ACCESSION AY532665  
 VERSION AY532665.1 GI:56462533  
 KEYWORDS .  
 SOURCE West Nile virus (WNV)  
 ORGANISM West Nile virus  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 Flavivirus; Japanese encephalitis virus group.  
 REFERENCE 1 (bases 1 to 11038)  
 AUTHORS Yamshchikov,G., Borisevich,V., Seregin,A., Chaporgina,E.,  
 Mishina,M., Mishin,V., Wai Kwok,C. and Yamshchikov,V.  
 TITLE An attenuated West Nile prototype virus is highly immunogenic and  
 protects against the deadly NY99 strain: a candidate for live WN  
 vaccine development  
 JOURNAL Virology 330 (1), 304-312 (2004)  
 PUBMED 15527855  
 REFERENCE 2 (bases 1 to 11038)  
 AUTHORS Borisevich,V.G. and Yamshchikov,V.F.  
 TITLE Molecular basis of attenuation of the West Nile virus prototype  
 strain B956  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 11038)  
 AUTHORS Borisevich,V.G. and Yamshchikov,V.F.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-JAN-2004) Molecular Biosciences, University of  
 Kansas, 1200 Sunnyside Ave., Lawrence, KS 66045, USA  
 FEATURES Location/Qualifiers  
 source 1. .11038  
 /organism="West Nile virus"  
 /mol\_type="genomic RNA"  
 /strain="B956"  
 /db\_xref="taxon:11082"  
 /note="obtained from R. Shope, Galveston, TX at suckling  
 mouse brain passage 2; passaged once in C6/36 cells"  
 CDS 97. .10389  
 /codon\_start=1  
 /product="polyprotein"

/protein\_id="AAT02759.1"  
/db\_xref="GI:56462534"  
/translation="MSKKPGPGKNAVNMLKRGMPRLSLIGLKRAMLSLIDGKGPI  
RFVLALLAFAFRFTAIPTRAVLDRWRGVNKQTAMKHLLSFKEELGT LTSAINRRSTKQ  
KKRGGTAGFTILLGLIACAGAVTLSNFQGVMMTVNATDVTDVITIPTAAGKNLCIVR  
AMDVGYLCEDTITYECPVLAAGNDPEDIDCWCTKSSVYVRYGRCTKTRHSRRSRSLT  
VQTHGESTLANKKGAWLDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV  
VFAILLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGSCVTLMSKDPTIDVK  
MMNMEAANLADVRSYCYLASVSDLSTRAACPTMGEAHNEKRADPAFVCKQGVVDRGWG  
NGCGLFGKGSIDTCAKFACTTKATGWIIQKENIKYEVAIFVHGPTTVEHKGIGATQA  
GRFSITPSAPSYTTLKGEYGEVTVDCEPRSGIDTSAYYMSVGAFLVHREWFMDLN  
LPWSSAGSTTWRNRETLMEFEEPHATKQSVVALGSQEGALHQALAGAIPEFSSNTVK  
LTSGHLKCRVCKMEKLQLKGTTYGVCSKAFKFARTPADTGHGVVLELQYTGTDGPCKV  
PISSVASLNLDTPVGRVTVNPVFSVATANSKVLIIELEPPFGDSYIVVGRGEQQINHH  
WHKSGSSIGKAFTTTLRGAQRЛАALGDTAWDFGSVGGVFTSVGKAIHQVFGAFRSLF  
GGMSWIQQLLGALLWMGINARDRSIAMTFLAVGGVLLFLSVNVHADTGCAIDIGRQ  
ELRCGSGVFIHNDVEAMDRYKFYPETPQGLAKIIQKAHAEGVCGLRSVSRLEHQMWE  
AIKDELNTLLKENGVDLSVVEKQNGMYKAAPKRLAATTEKLEMGWKAWGKSIIFAPE  
LANNTFVIDGPETEECPTANRAWNSMEVEDFGFGLTSTRMFLRIRETNTTECDSKIIG  
TAVKNNMAVHSDSLWIESGLNDTWKLERAVLGEVKSCSWPETHLWGDRVLESIDLII  
PITLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTISDSCGHRGPAARTT  
TESGKLITDWCCRSCPLPLRFQTEENGWYGMEIRPTRHDEKTLVQSRVNAYNADMID  
PFQLGLVVFLATQEVLRKRWTAKISI PAIMLALLVLVFGGITYTDVLRYVILVGAAF  
AEANSGGDVVHLALMATFKIOPVFLVASFLKARWTNQESILLMLAAFFQMAYYDAKN  
VLSWEVPDVNLNSLVAWMILRAISFTNTSNVVVPLLALLTPGLKCLNLDVYRILLMV  
GVGSLIKEKRSSAAKKGACLICLALASTGVFPNPMILAAGLMACDPNRKRGWPATEVM  
TAVGLMFIAVGGLAELDIDSMAIPMTIAGLMFVAFVISGKSTDWIERTADITWESDA  
EITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVIGFWI  
TLQYTKRGGVWLWDTSPKEYKKGDTTGVRIMTRGLLGSYQAGAGVMVEGFHTLWH  
TTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNHDEVQMI VVEPGKVN  
QTKPGVFKTPEGEIGAVTLDYPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYISAIVQ  
GERMEEPAPAGFEPEMLRKQITVLDLHPGAGKTRKILPQIIKEAINKRLRTAVLAPT  
RVVAAEMSEALRGLPIRYQTSAVHREHSGNEIVDVMCHATLTHRLMSPHRVPNYNLF  
MDEAHFTDPASIAARGYIATKVELGEAAAFMTATPPGTSDFPESNAPISDMQTEIP  
DRAWNTGYEWITEYVGTVWFPSVKGNEIALCLQRAGKKVIQLNRKSYETEYPKCK  
NDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIIIEGDGRVILGEPSAITAASAAQ  
RRGRIGRNPQVGDECYGGHTNEDDSNFAHWTEARIMLDNINMPNGLVAQLYQPERE  
KVTMDGEYRLRGEERKNFLEFLRTADLPWLAYKVAAGISYHDRKWCDFGPRNTI  
LEDNNEVEVITKLGKERRKILRPRWADARVYSDHQALKSFKDFASGKRSQIGLVEVLGRM  
PEHFMGKTWEALDTMYVATAEKGGRAHRMALEELPDALQTIALLSVMMSLGVFFL  
LMQRKGIGKIGLGGIILGAATFFCWMAEVPGTKIAGMLLSLLMIVLIPPEPEKQRSQ  
TDNQLAVFLICVLTGAVAANEMGWLDKTKNDISSLGHKPEARETTLGVESFLDL  
RPATAWSLYAVTTAVLTPLLKHLITSODYINTSLTSINVQASALFTLARGFPFVDVGVS  
ALLLAAGCWQVTLTVTAAALLFCHYAYMVPGWQAEAMRSAQRTAACIMKNVVWD  
GIVATDVELERTTPVMQKKVQII LILVSMMAVVNVPSVRTVREAGILTTAAAVTLW  
ENGASSVWNATTAGLCHIMRGGLSCLSIMWTLIKNMEKPLKRGGAKGRTLGEVWK  
ERLNHMTKEEFTTRYREAITEVDRSAAKHARREGNITGGHPVSRGTAKLRWLVERRFL  
EPVGKVVDLGCGRGWCYYMATQKRVQEVKGTYKGGPGEHEPQLVQSYGWNIVTMKSG  
VDVFYRPSEASDTLLCDIGESSSSAEVEEHTRVLEMVEDWLHGRPKEFCIKVLCPY  
MPKVIEKMETLQRRYGGGLVRNPLSRNSTHEMYWVSHASGNIVHSVNMTSQVLLGRME  
KKTWKGPQFEEVDNLGGSTRAVGKPLLNSDTSKIKNRIERLKKEYSSTWHQDANHPYR  
TWNYHGSYEVKPTGSASSLVNGVVRLLSKPWDITNVTTMAMTDTPFGQQRVFKEV  
DTKAPEPPEGVKYVLNNTNWLAFLARDKKPRMCSREEFIGKVN SNAALGAMFEEQN  
QWKNAREAVEDPKFWEMVDEEREAHLRGECNTCIYNNMMGKREKKPGEFGKAKGSRAIW  
FMWL GARFLEFA LGFLNEDH WLGRKNSGGVEGLGLQKLG YILKEVGT KPGGKVYAD  
DTAGWDTRITKADLEN EAKVLELLDGEH RRLARSIIELTYRH KVVKVMRPAADGKTV  
DVISREDQRGSGQVVTYALNTFTNLAVQLVRRMMEGEVGIGPDDVEKL GKGKGPKVRTW

L F E N G E E R L S R M A V G D D C V V K P L D D R F A T S L H F L N A M S K V R K D I Q E W K P S T G W Y D W Q  
Q V P F C S N H F T E L I M K D G R T L V V P C R G Q D E L I G R A R I S P G A G W N V R D T A C L A K S Y A Q M W  
L L L Y F H R R D L R L M A N A I C S A V P A N W P T G R T T W S I H A K G E W M T T E D M L A V W N R V W I E E  
N E W M E D K T P V E R W S D V P Y S G K R E D I W C G S L I G T R T R A T W A E N I H V A I N Q V R S V I G E E K  
Y V D Y M S S L R R Y E D T I V V E D T V L "

## ORIGIN

Query Match 86.9%; Score 1321; DB 10; Length 11038;  
Best Local Similarity 92.6%; Pred. No. 0;  
Matches 1403; Conservative 0; Mismatches 100; Indels 12; Gaps 1;

Qy	1	CGGAATTCACTGTTAGGAATGAGCAACAGGGACTCCTGGAGGGAGTGTCTG	60
Db	956	CAGCATACAGCTCACTGCTTAGGAATGAGTAACAGAGACTCCTGGAGGGAGTGTCTG	1015
Qy	61	GAGCTACATGGGTTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAAG	120
Db	1016	GAGCTACATGGGTTGATCTGGTACTGGAAGGCGATAGTTGTGTGACCCTAATGTCAAAAG	1075
Qy	121	ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAATCTCGCAGATGTGC	180
Db	1076	ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCCAACCTCGCAGATGTGC	1135
Qy	181	GTAGCTACTGCTACTTAGCTCGGTCACTGATCTGTCAACAAAAGCCCGTGTCCAACCA	240
Db	1136	GCAGTTACTGTTACCTAGCTCGGTCACTGACTTGTCAACAAAGAGCTGCGTGTCCAACCA	1195
Qy	241	TGGGTGAAGCTCACAAACGAGAAAAGAGCCGACCCCTGCCCTTGTGCAAGCAAGGCGTCG	300
Db	1196	TGGGTGAAGCCCACAACGAGAAAAGAGCTGACCCGCCCTCGTTGCAAGCAAGGCGTTG	1255
Qy	301	TAGACAGAGGATGGGAAATGGATGCGGACTGTTGGAAAGGGGAGCATTGACACATGTG	360
Db	1256	TGGACAGAGGATGGGAAATGGCTGCGGACTGTTGGAAAGGGGAGCATTGACACATGTG	1315
Qy	361	CAAAGTTGCCTGTACAACCAAGGCAACTGGTTGGATTATCAGAAGGAAAACATCAAGT	420
Db	1316	CGAACAGTTGCCTGTACAACCAAGCAACTGGATGGATCATCCAGAAGGAAAACATCAAGT	1375
Qy	421	ACGAGGTTGCCATATTGTGCATGGCCGACGACTGTCGAATCACATGGCAATTATTCAA	480
Db	1376	ATGAGGTTGCCATATTGTGCATGGCCGACGACCGTTGAATCTCATGGCA-----	1426
Qy	481	CACAGATAGGGCTACCCAAGCAGGAAGGTTCAGCATAACTCCATGGCACCATCCTACA	540
Db	1427	---AGATAGGGCCACCCAGGCTGGAAGATTCACTCCATGGCACCATCTTACA	1483
Qy	541	CGCTGAAGTTGGTGAGTATGGTGAGGTACAGTTGACTGTGAGCCACGGTCAGGAATAG	600
Db	1484	CGCTAAAGTTGGTGAGTATGGTGAGGTACGGTTACGGTTGATTGTGAGCCACGGTCAGGAATAG	1543
Qy	601	ACACTAGCGCTTACTACGTTATGTCACTGGTGAGGTACAGTTGACTGTGAGCCACGGTCAGGAAT	660
Db	1544	ACACTAGCGCCTATTACGTTATGTCACTGGTGAGGTACAGTCCTGGTTACCGGAGAAT	1603
Qy	661	GGTTTATGGACCTGAACCTCCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG	720

Db 1604 GGTTTATGGATCTGAACCTGCCATGGAGCAGTGCTGGAAGCACCGACCACGTGGAGGAACCGGG 1663  
Qy 721 AAACACTGATGGAGTTGAAGAACCTCATGCCACCAAAACAATCTGTCGTAGCTCTAGGGT 780  
|||  
Db 1664 AAACACTGATGGAGTTGAAGAACCTCATGCCACCAAAACAATCTGTTGTGGCTCTAGGGT 1723  
Qy 781 CGCAGGAAGGTGCCTTGCACCAAGCTCTGGCTGGAGCAATTCTGTTGAGTTCTCAAGCA 840  
|||  
Db 1724 CGCAGGAAGGTGCCTTGCACCAAGCTCTGGCCGGAGCGATTCTGTTGAGTTCTCAAGCA 1783  
Qy 841 ACACGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTAAGATGGAGAAGTTGCAGC 900  
|||  
Db 1784 ACACGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTAAGATGGAGAAGTTGCAGC 1843  
Qy 901 TGAAGGAAACAACATATGGTGTATGCTAAAAGCATTCAAATTGCTAGGACTCCGCTG 960  
|||  
Db 1844 TGAAGGAAACAACATATGGAGTATGTTCAAAGCGTTCAAATTGCTAGGACTCCGCTG 1903  
Qy 961 ACACGGTCATGGAACGGTGGTGTGGAACTGCAGTATACCGGAAAAGACGGGCCTTGCA 1020  
|||  
Db 1904 ACACGGCCACGGAACGGTGGTGTGGAACTGCAATATACCGGAAACAGACGGTCCCTGCA 1963  
Qy 1021 AAGTGCCCATTCCTCTGTGGCTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGA 1080  
|||  
Db 1964 AAGTGCCCATTCCTCCGTAGCTCCCTGAATGACCTCACACCTGTTGGAAGACTGGTGA 2023  
Qy 1081 CTGTGAATCCATTGTGTCTGTGGCTACGGCAACTCGAAGGTTTGATTGAACCTCGAAC 1140  
|  
Db 2024 CCGTGAATCCATTGTGTCTGTGGCCACAGCCAACCTCGAAGGTTTGATTGAACCTCGAAC 2083  
Qy 1141 CCCCGTTAGTGACTCTTACATCGTGGGGAGAGGAGAACAGCAGATAAACCAACACT 1200  
|||  
Db 2084 CCCCGTTGGTGAACCTTACATCGTGGGGAGAGGAGAACAGCAGATAAACCATCACT 2143  
Qy 1201 GGCACAAATCTGGAGCAGTATTGAAAGGCTTCACCACTACACTCAGAGGAGCTAAC 1260  
|||  
Db 2144 GGCACAAATCTGGAGCAGCATTGAAAGGCCTTACCAACACTCAGAGGAGCTAAC 2203  
Qy 1261 GACTTGCAGCTTGGAGACACTGCCTGGATTTGGATCAGTCGGAGGGTTTCACCT 1320  
|||  
Db 2204 GACTCGCAGCTTGGAGATACTGCTTGGATTTGGATCAGTTGGAGGGTTTCACCT 2263  
Qy 1321 CGTAGGGAAAGCCATACACCAAGTTTGGAGGAGCCTTAGATCACTTTGGAGGGA 1380  
|  
Db 2264 CAGTGGGGAAAGCCATACACCAAGTCTTGGAGGAGCTTAGATCACTTTGGAGGGA 2323  
Qy 1381 TGTCTGGATCACACAGGGCTTCTGGAGCTTCTGCTGTGGATGGAATTACGCC 1440  
|||  
Db 2324 TGTCTGGATCACACAGGGACTTCTGGAGCTTCTGTTGTGGATGGAATTCAATGCC 2383  
Qy 1441 GTGACAGGTCAATTGCTATGACGTTCTTGCAGGGTGGAGGAGTCTGCTCTCCTTCGG 1500  
|||  
Db 2384 GTGACAGGTCAATTGCTATGACGTTCTTGCAGGGTGGAGGAGTTTGCTCTCCTTCGG 2443  
Qy 1501 TCAACGTCCATGCTG 1515  
|||  
Db 2444 TCAACGTCCATGCTG 2458

## RESULT 5

DQ318019

LOCUS DQ318019 11038 bp mRNA linear VRL 01-JAN-2006

DEFINITION West Nile virus strain ArD76104, complete genome.

ACCESSION DQ318019

VERSION DQ318019.1 GI:84028432

KEYWORDS .

SOURCE West Nile virus (WNV)

ORGANISM West Nile virus

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.

REFERENCE 1 (bases 1 to 11038)

AUTHORS Borisevich,V.G., Seregin,A.V. and Yamshchikov,V.F.

TITLE Genetic determinants of West Nile virus pathogenicity

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 11038)

AUTHORS Borisevich,V.G. and Yamshchikov,V.F.

TITLE Direct Submission

JOURNAL Submitted (07-DEC-2005) Molecular Biosciences, 1200 Sunnyside ave, Lawrence, KS 66045, USA

FEATURES Location/Qualifiers

source 1. .11038  
           /organism="West Nile virus"  
           /mol\_type="mRNA"  
           /strain="ArD76104"  
           /db\_xref="taxon:11082"  
           /country="Senegal"  
           /note="lineage 2; SMB pass 3, C6/36 pass 1"

5'UTR 1. .96  
        97. .10389

CDS /codon\_start=1  
       /product="polyprotein"  
       /protein\_id="ABC49716.1"  
       /db\_xref="GI:84028433"  
       /translation="MSKKPGGPGKNRAVNMLKRGMPRLSLIGLKRAMLSLIDGKGPI  
           RFVLALLAFFFRTAIAPTRAVLDRWRCVNQTA  
           MKHLLSFKKELGTLSAINRRSTKQ  
           KKRGGTAGFTILLGLIACAGAVTLSNFQGKVMMTVNATDV  
           TDVITIPTAA  
           GKNLCIVR  
           AMDVGYLCEDTITYECPVLAAGNDPEDI  
           DCWCTKSSVVYR  
           GRCTKTRHSRRSRSRSLT  
           VQTHGESTLANKGA  
           WL  
           DSTKATRYLV  
           KTESWILRNPGYALVA  
           VAVIGWMLGSNTMQRV  
           VFAI  
           LLL  
           VAPAYSFNCLGMSNRDFLEG  
           VGATWVDL  
           LEGD  
           CVTIMSKDKPTIDVK  
           MMNMEAANLAD  
           VRSYCYLASV  
           SDL  
           TRAAC  
           PTMGEAHNE  
           KRADPAF  
           VCKQGV  
           VDRGWG  
           NGCGLFGKGSID  
           TCAKFACT  
           TTKATGWI  
           IQKENI  
           KYEV  
           AFVHG  
           P  
           TTV  
           VES  
           HKG  
           KIG  
           ATQA  
           GRFSITPSAPS  
           YT  
           TLK  
           LGEY  
           GEV  
           TV  
           DCE  
           PRSG  
           IDTS  
           SAY  
           YV  
           MSVG  
           AKS  
           FLV  
           HREW  
           FMDLN  
           LPW  
           SSAG  
           ST  
           WRN  
           RET  
           LVE  
           FE  
           EPH  
           AT  
           KQS  
           V  
           VAL  
           LG  
           SQEG  
           AL  
           HQ  
           AL  
           AG  
           AI  
           P  
           VE  
           F  
           SS  
           NT  
           V  
           T  
           V  
           L  
           Q  
           Y  
           T  
           G  
           D  
           P  
           C  
           K  
           V  
           I  
           S  
           V  
           A  
           N  
           D  
           L  
           T  
           P  
           V  
           G  
           R  
           L  
           T  
           V  
           N  
           P  
           F  
           V  
           S  
           V  
           A  
           T  
           A  
           N  
           K  
           V  
           L  
           I  
           E  
           L  
           P  
           P  
           F  
           G  
           D  
           S  
           Y  
           I  
           V  
           V  
           G  
           R  
           G  
           E  
           Q  
           Q  
           I  
           N  
           H  
           H  
           W  
           H  
           K  
           S  
           G  
           S  
           I  
           G  
           K  
           A  
           F  
           T  
           T  
           T  
           L  
           R  
           G  
           A  
           Q  
           R  
           L  
           A  
           L  
           G  
           D  
           T  
           A  
           W  
           F  
           G  
           S  
           V  
           G  
           V  
           F  
           T  
           S  
           V  
           G  
           K  
           A  
           I  
           H  
           Q  
           V  
           F  
           G  
           A  
           F  
           R  
           S  
           L  
           F  
           G  
           M  
           S  
           W  
           I  
           Q  
           G  
           L  
           G  
           A  
           L  
           R  
           M  
           I  
           A  
           T  
           F  
           L  
           A  
           G  
           V  
           G  
           F  
           V  
           I  
           G  
           T  
           Y  
           P  
           E  
           R  
           T  
           H  
           D  
           E  
           K  
           T  
           L  
           V  
           Q  
           S  
           R  
           V  
           N  
           A  
           Y  
           A  
           D  
           M  
           I  
           D  
           P  
           F  
           Q  
           L  
           G  
           L  
           L  
           V  
           V  
           F  
           L  
           A  
           T  
           Q  
           E  
           V  
           L  
           R  
           K  
           R  
           W  
           T  
           A  
           K  
           I  
           S  
           I  
           P  
           A  
           I  
           M  
           A  
           L  
           L  
           V  
           L  
           V  
           F  
           G  
           G  
           I  
           T  
           Y  
           D  
           V  
           L  
           R  
           V  
           I  
           L  
           G  
           A  
           F

AEANSGGDVVHLALMATFKIQPVFLVASFLKARWTNQESIILMLAAAFQOMAYYDAKN  
VLSWEVPDVLSLSVAWMILRAISFTNTSNVVVPLLALLTPGLKCLNLDVYRILLMV  
GVGSLIKEKRSSAAKKGACLICLALASTGVFNPMLAAGLMACDPNRKGWPATEVM  
TAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFVAFVISGKSTDMWIERTADITWESDA  
EITGSSERVDVRLDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVIGFWI  
TLQYTKRGGVWLWDTSPKEYKKGDTTGVYRIMTRGLLGSYQAGAGVMVEGFHTLWH  
TTKGAAALMSGEGRLLDPYWGVSKEEDRLCYGGPWKLQHKGWNHGDEVQMI  
VVEPGKVNKNVQTKPGVFKTPEGEIGAVTLDYPTGTSGSPIVDKNGDVGILYGN  
GNGVIMPNGSYISAIVQGERMEEPAPAGFEPEMLRKQITVLDLHPGAGKTRK  
ILPQIKEAINKRLRTAVLAPT  
RVVAAEMSEALRGLPIRYQTSAVHREHSGNEIVDMCHATLTHRLMSPHRVP  
NYNLFI  
MDEAHFTDPASIAARGYIATKVELGEAAAI  
FMTATPPGTSDPF  
PESNAPI  
SDMQTEIP  
DRAWNTGYEWITEYVGKTVWFVPSVKMGNEIALCLQRAGKKV  
IQLNRKS  
YETEYPKCK  
NDDWDFVITTDISEMGANFKASRVIDSRKSVKPT  
IIEEGDGRVILGEPS  
AITAASAAQ  
RRGRIGRNP  
SQVGDECYGGHT  
NEDDSNFAH  
WTEARIMLDN  
INMPNGLVAQ  
LYQP  
ERE  
KVYTM  
DGEYR  
LGEER  
KNF  
LEFL  
RTADLP  
VWL  
AYK  
VAAAG  
ISY  
HDR  
KWC  
FDG  
PRT  
NTI  
LEDN  
NEVE  
VIT  
KL  
GER  
KIL  
LPR  
RWAD  
ARV  
YSD  
HQAL  
KSFK  
DAS  
GKRS  
SQIG  
LVE  
VLGRM  
PEH  
FMGK  
TWEA  
LDTM  
VVATA  
EK  
GGRA  
HRM  
AEL  
LP  
DAL  
QT  
IAL  
L  
S  
VMS  
LG  
VFFL  
LMQR  
KG  
IG  
K  
I  
LG  
GV  
I  
LGA  
AT  
FFC  
WMA  
EV  
PGT  
K  
IAG  
M  
L  
L  
S  
L  
L  
M  
I  
V  
L  
I  
PE  
PE  
K  
QR  
SQ  
TDN  
QLA  
VFL  
IC  
VLT  
L  
V  
SA  
A  
N  
EM  
G  
W  
L  
D  
K  
T  
K  
N  
D  
I  
G  
S  
L  
L  
G  
H  
K  
P  
E  
A  
R  
T  
T  
L  
G  
V  
E  
S  
F  
L  
D  
L  
R  
P  
A  
T  
A  
W  
S  
L  
Y  
A  
V  
T  
T  
A  
V  
L  
T  
P  
L  
L  
K  
H  
L  
I  
T  
S  
D  
Y  
I  
N  
T  
S  
L  
T  
S  
I  
N  
V  
Q  
A  
S  
A  
L  
F  
T  
L  
A  
R  
G  
F  
P  
V  
D  
V  
G  
V  
S  
A  
L  
L  
A  
A  
G  
C  
W  
Q  
V  
T  
L  
T  
V  
T  
A  
A  
L  
L  
F  
C  
H  
Y  
A  
M  
V  
P  
G  
W  
Q  
A  
E  
A  
M  
R  
S  
A  
Q  
R  
R  
T  
A  
A  
G  
I  
M  
K  
N  
A  
V  
V  
D  
G  
I  
V  
A  
T  
D  
V  
P  
E  
L  
E  
R  
T  
T  
P  
V  
M  
Q  
K  
V  
G  
Q  
I  
M  
L  
I  
L  
V  
S  
M  
A  
V  
V  
V  
N  
P  
S  
V  
R  
T  
V  
R  
E  
A  
G  
I  
L  
T  
T  
A  
A  
V  
T  
L  
W  
E  
N  
G  
A  
S  
S  
V  
W  
N  
A  
T  
A  
I  
G  
L  
C  
H  
I  
M  
R  
G  
G  
W  
L  
S  
C  
L  
S  
I  
M  
W  
T  
L  
I  
K  
N  
M  
E  
K  
P  
G  
L  
K  
R  
G  
G  
A  
K  
G  
R  
T  
L  
G  
E  
V  
W  
K  
E  
R  
L  
N  
H  
M  
T  
K  
E  
E  
F  
T  
R  
Y  
R  
K  
E  
A  
I  
T  
E  
V  
D  
R  
S  
A  
A  
K  
H  
A  
R  
R  
E  
G  
N  
I  
T  
G  
G  
H  
P  
V  
S  
R  
G  
T  
A  
K  
L  
R  
W  
L  
V  
E  
R  
F  
L  
E  
P  
V  
G  
K  
V  
V  
D  
L  
G  
C  
G  
R  
G  
G  
W  
C  
Y  
Y  
M  
A  
T  
Q  
K  
R  
V  
Q  
E  
V  
K  
G  
Y  
T  
K  
G  
G  
P  
H  
E  
E  
P  
Q  
L  
V  
Q  
S  
Y  
G  
W  
N  
I  
V  
T  
M  
K  
S  
G  
V  
D  
V  
F  
Y  
R  
P  
S  
E  
A  
S  
D  
T  
L  
L  
C  
D  
I  
G  
E  
S  
S  
S  
A  
E  
V  
E  
E  
H  
R  
T  
V  
R  
V  
L  
E  
M  
V  
E  
D  
W  
L  
H  
R  
G  
P  
K  
E  
F  
C  
I  
K  
V  
L  
C  
P  
Y  
M  
P  
K  
V  
I  
E  
K  
M  
E  
T  
L  
Q  
R  
R  
Y  
G  
G  
G  
L  
L  
V  
R  
N  
P  
L  
S  
R  
N  
S  
T  
H  
E  
M  
Y  
W  
V  
S  
H  
A  
G  
N  
I  
V  
H  
S  
V  
N  
M  
T  
S  
Q  
V  
L  
L  
G  
R  
M  
E  
K  
K  
T  
W  
K  
G  
P  
Q  
F  
E  
E  
D  
V  
N  
L  
G  
S  
G  
T  
R  
A  
V  
G  
K  
P  
L  
L  
N  
S  
D  
T  
S  
K  
I  
K  
N  
R  
I  
E  
R  
L  
K  
E  
Y  
S  
S  
T  
W  
H  
Q  
D  
A  
N  
H  
P  
Y  
R  
T  
W  
N  
Y  
H  
G  
S  
Y  
E  
V  
K  
P  
T  
G  
S  
A  
S  
L  
V  
N  
G  
V  
V  
R  
L  
L  
S  
K  
P  
W  
D  
T  
I  
T  
N  
V  
T  
T  
M  
A  
M  
T  
D  
T  
P  
F  
G  
Q  
Q  
R  
V  
F  
K  
E  
K  
V  
D  
T  
K  
A  
P  
E  
P  
P  
E  
G  
V  
K  
Y  
V  
L  
N  
E  
T  
T  
N  
W  
L  
W  
A  
F  
L  
A  
R  
D  
K  
K  
P  
R  
M  
C  
S  
R  
E  
E  
F  
I  
G  
V  
K  
V  
N  
S  
N  
A  
L  
G  
A  
M  
F  
E  
Q  
N  
Q  
W  
K  
N  
A  
R  
E  
A  
V  
D  
P  
K  
F  
W  
E  
M  
V  
D  
E  
E  
R  
A  
H  
L  
R  
G  
E  
C  
N  
T  
C  
I  
Y  
N  
M  
M  
G  
K  
R  
E  
K  
K  
P  
G  
E  
F  
G  
K  
A  
G  
S  
R  
A  
I  
W  
F  
M  
W  
L  
G  
A  
R  
F  
L  
E  
F  
A  
L  
G  
F  
L  
N  
E  
D  
H  
W  
L  
G  
R  
K  
N  
S  
G  
G  
V  
E  
G  
L  
G  
L  
Q  
K  
L  
G  
Y  
I  
L  
K  
E  
V  
G  
T  
K  
P  
G  
G  
K  
V  
Y  
A  
D  
D  
T  
A  
G  
W  
D  
T  
R  
I  
T  
K  
A  
D  
L  
E  
N  
E  
A  
K  
V  
L  
E  
L  
L  
D  
G  
E  
H  
R  
R  
L  
A  
R  
S  
I  
I  
E  
L  
T  
Y  
R  
K  
V  
V  
K  
M  
R  
P  
A  
A  
D  
G  
K  
T  
V  
M  
D  
V  
I  
S  
R  
E  
D  
Q  
R  
G  
S  
G  
Q  
V  
V  
T  
Y  
A  
L  
N  
T  
F  
T  
N  
L  
A  
V  
Q  
L  
V  
R  
M  
M  
E  
G  
E  
G  
V  
I  
G  
P  
D  
V  
E  
K  
L  
G  
K  
G  
P  
K  
V  
R  
T  
W  
L  
F  
E  
N  
G  
E  
E  
R  
L  
S  
R  
M  
A  
V  
G  
S  
D  
D  
C  
V  
V  
K  
P  
L  
D  
D  
R  
F  
A  
T  
S  
L  
H  
F  
L  
N  
A  
M  
S  
K  
V  
R  
K  
D  
I  
Q  
E  
W  
K  
P  
S  
T  
G  
W  
Y  
D  
W  
Q  
V  
P  
F  
C  
S  
N  
H  
F  
T  
E  
L  
I  
M  
K  
D  
G  
R  
T  
L  
V  
V  
P  
C  
R  
G  
Q  
D  
E  
L  
I  
G  
R  
A  
R  
I  
S  
P  
G  
A  
G  
N  
V  
R  
D  
T  
A  
C  
L  
A  
K  
S  
Y  
A  
Q  
M  
W  
L  
L  
L  
Y  
F  
H  
R  
R  
D  
L  
R  
L  
M  
A  
N  
A  
I  
C  
S  
A  
V  
P  
V  
N  
W  
P  
T  
G  
R  
T  
T  
W  
S  
I  
H  
A  
K  
G  
E  
W  
M  
T  
T  
E  
D  
M  
L  
A  
V  
W  
N  
R  
V  
W  
I  
E  
E  
N  
W  
E  
M  
D  
K  
T  
P  
V  
E  
R  
W  
S  
D  
V  
P  
Y  
S  
G  
K  
R  
E  
D  
I  
W  
C  
G  
S  
L  
I  
G  
T  
R  
T  
R  
A  
T  
W  
A  
E  
N  
I  
H  
V  
A  
I  
N  
Q  
V  
R  
S  
V  
I  
G  
E  
E  
K  
K  
Y  
D  
Y  
M  
S  
S  
L  
R  
R  
Y  
E  
D  
T  
I  
V  
V  
E  
D  
T  
V  
L  
"

```
mat_peptide    97. .411
                /product="C protein"
sig_peptide   412. .465
                /note="prM signal peptide"
mat_peptide   466. .966
                /product="prM protein"
mat_peptide   466. .741
                /product="cleaved amino terminal prM fragment"
mat_peptide   742. .966
                /product="M protein"
sig_peptide   919. .966
                /note="E signal peptide"
mat_peptide   967. .2457
                /product="E protein"
sig_peptide   2386. .2457
                /note="NS1 signal peptide"
mat_peptide   2458. .3513
                /product="NS1 protein"
```

mat_peptide	3514.	.4206
		/product="NS2A protein"
mat_peptide	4207.	.4599
		/product="NS2B protein"
mat_peptide	4600.	.6456
		/product="NS3 protein"
mat_peptide	6457.	.6903
		/product="NS4A protein"
sig_peptide	6835.	.6903
		/note="2K peptide"
mat_peptide	6904.	.7671
		/product="NS4B protein"
mat_peptide	7672.	.10386
		/product="NS5 protein"
3'UTR	10390.	.11038

#### ORIGIN

Query Match                    86.9%; Score 1321; DB 10; Length 11038;  
 Best Local Similarity    92.6%; Pred. No. 0;  
 Matches 1403; Conservative 0; Mismatches 100; Indels 12; Gaps 1;

Qy	1	CGGAATTCACTGTTAGGAATGAGCAACAGGGACTTCCTGGAGGGAGTGTCTG	60
Db	956	CAGCATAACAGCTTCACTGCTTAGGAATGAGAACAGAGACTTCCTGGAGGGAGTGTCTG	1015
Qy	61	GAGCTACATGGGTTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTC AAAAG	120
Db	1016	GAGCTACATGGGTTGATCTGGTACTGGAAGGCAGATGTTGTGTGACCATAATGTC AAAAG	1075
Qy	121	ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAATCTCG CAGATGTGC	180
Db	1076	ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCCAACCTCG CAGATGTGC	1135
Qy	181	GTA GCTACTGCTACTTAGCTCGTCAGTGATCTGTCAACAAAAGCC CGTGTCCAACCA	240
Db	1136	GCAGTTACTGTTACCTAGCTCGTCAGTGACTTGTCAACAAGAGCT CGTGTCCAACCA	1195
Qy	241	TGGGTGAAGCTCACACGAGAAAAGAGCCGACCC CTGCCTTGTGCAAGCAAGGCGTCG	300
Db	1196	TGGGTGAAGCCCACACGAGAAAAGAGCT GACCCCGCCTCGTTGCAAGCAAGGCGTTG	1255
Qy	301	TAGACAGAGGATGGGAAATGGAT GCGGACTGTTGAAAGGGAGCATTGACACATGTG	360
Db	1256	TGGACAGAGGATGGGAAATGGCT GCGGACTGTTGAAAGGGAGCATTGACACATGTG	1315
Qy	361	CAAAGTTGCC TGTACAACCAAGGCAACTGGTTGGATTATCC AGAAGGAAAACATCAAGT	420
Db	1316	CGAAGTTGCC TGTACAACCAAAGCAACTGGATGGATCAT CCAGAAGGAAAACATCAAGT	1375
Qy	421	ACGAGGTTGCC CATATTGTGCATGGCCCGACGACT GTGCAATCACATGGCAATTATTCAA	480
Db	1376	ATGAGGTTGCC CATATTGTGCATGGCCCGACGACCGTTGAAT CTCATGGCA-----	1426
Qy	481	CACAGATAGGGCTACCC AAGCAGGAAGGTT CAGCATAACTCC CATCGGCACC CATCCTACA	540
Db	1427	---AGATAGGGCCACCC AAGGCTGGAAGATT CAGTATAACTCC CATCGGCAC CATCTTACA	1483

Qy	541	CGCTGAAGTTGGGTGAGTATGGTGAGGTACAGTTGACTGTGAGCCACGGTCAGGAATAG	600
Db	1484	CGCTAAAGTTGGGTGAGTATGGTGAGGTTACGGTGATTGTGAGCCACGGTCAGGAATAG	1543
Qy	601	ACACTAGCGCTTACTACGTTATGTCAGTGGTGCAGAAGTCCTTGGTCACCGAGAAT	660
Db	1544	ACACTAGCGCTTACTACGTTATGTCAGTGGTGCAGAAGTCCTTGGTCACCGAGAAT	1603
Qy	661	GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG	720
Db	1604	GGTTTATGGATCTGAACCTGCCATGGAGCAGTGCTGGAAGCACCACGTGGAGGAACCGGG	1663
Qy	721	AAACACTGATGGAGTTGAAGAACCTCATGCCACCAAACAATCTGTCGTAGCTCTAGGGT	780
Db	1664	AAACACTGGTGGAGTTGAAGAACCTCATGCCACCAAACAATCTGTTGGCTCTAGGGT	1723
Qy	781	CGCAGGAAGGTGCCTTGCACCAAGCTCTGGCTGGAGCAATTCTGTTGAGTTCTCAAGCA	840
Db	1724	CGCAGGAAGGTGCCTTGCACCAAGCTCTGGCCGGAGCGATTCTGTTGAGTTCTCAAGCA	1783
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	900
Db	1784	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	1843
Qy	901	TGAAGGGAAACAACATATGGTGTATGCTAAAAGCATTCAAATTGCTAGGACTCCCGCTG	960
Db	1844	TGAAGGGAAACAACATATGGAGTATGTTCAAAGCGTTCAAATTGCTAGGACTCCCGCTG	1903
Qy	961	ACACTGGTCATGGAACCGTGGTGCCTGGAACACTGCAGTATACCGGAAAGACGGGCTTGCA	1020
Db	1904	ACACTGGCACGGAACCGTGGTGTGGAACTGCAATATAACCGGAACAGACGGTCCCTGCA	1963
Qy	1021	AAGTGCCATTCTCTGTGGCTCCCTGAACGACCTACACCCGTGGAAGGCTGGTGA	1080
Db	1964	AAGTGCCATTCTCCGTAGCTCCCTGAATGACCTCACACCTGTTGGAAGACTGGTGA	2023
Qy	1081	CTGTGAATCCATTGTGTCGTGGCTACGCCACTCGAAGGTTTGATTGAACCTCGAAC	1140
Db	2024	CCGTGAATCCATTGTGTCGTGGCCACAGCCAACCTCGAAGGTTTGATTGAACCTCGAAC	2083
Qy	1141	CCCCGTTAGTGAACCTTACATCGTGGTGGGAGAGGAGAACAGCAGATAAACCAACT	1200
Db	2084	CCCCGTTGGTGAACCTTACATCGTGGTGGGAAGAGGAGAACAGCAGATAAACCAACT	2143
Qy	1201	GGCACAAATCTGGGAGCAGTATTGAAAGGCTTCACCACTACACTCAGAGGAGCTAAC	1260
Db	2144	GGCACAAATCTGGGAGCAGCATTGAAAGGCTTCACCACTACACTCAGAGGAGCTAAC	2203
Qy	1261	GACTTGCAGCTCTGGAGACACTGCCTGGATTTGGATCAGTCGGAGGGTTTCACCT	1320
Db	2204	GACTCGCAGCTCTGGAGATACTGCTTGGATTTGGATCAGTTGGAGGGTTTCACCT	2263
Qy	1321	CGGTAGGGAAAGCCATACACCAAGTTTGGAGGAGCCTTAGATCACTCTTGGAGGGA	1380
Db	2264	CAGTGGGAAAGCCATACACCAAGTCTTGGAGGAGCTTAGATCACTCTTGGAGGGA	2323
Qy	1381	TGTCCTGGATCACACAGGGCTCTGGAGCTCTCTGCTGTGGATGGAAATTACGCC	1440

Db	2324	TGTCCTGGATCACACAGGGACTTCTGGGAGCTCTCTGTGGATGGGAATCAATGCC	2383
Qy	1441	GTGACAGGTCAATTGCTATGACGTTCTTGC GGTTGGAGGAGTCTTGCTCTTCCTTCGG	1500
Db	2384	GTGACAGGTCAATTGCTATGACGTTCTTGC GGTTGGAGGAGTTTGCTCTTCCTTCGG	2443
Qy	1501	TCAACGTCCATGCTG	1515
Db	2444	TCAACGTCCATGCTG	2458

## RESULT 6

WNFCG  
 LOCUS WNFCG 10962 bp ss-RNA linear VRL 08-MAY-2002  
 DEFINITION West Nile virus RNA, complete genome.  
 ACCESSION M12294 M10103  
 VERSION M12294.2 GI:11497619  
 KEYWORDS .  
 SOURCE West Nile virus  
 ORGANISM West Nile virus  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 Flavivirus; Japanese encephalitis virus group.  
 REFERENCE 1 (bases 67 to 969)  
 AUTHORS Castle,E., Nowak,T., Leidner,U., Wengler,G. and Wengler,G.  
 TITLE Sequence analysis of the viral core protein and the  
 membrane-associated proteins V1 and NV2 of the flavivirus West Nile  
 virus and of the genome sequence for these proteins  
 JOURNAL Virology 145 (2), 227-236 (1985)  
 PUBMED 2992152  
 REFERENCE 2 (bases 859 to 2658)  
 AUTHORS Wengler,G., Castle,E., Leidner,U., Nowak,T. and Wengler,G.  
 TITLE Sequence analysis of the membrane protein V3 of the flavivirus West  
 Nile virus and of its gene  
 JOURNAL Virology 147 (2), 264-274 (1985)  
 PUBMED 3855247  
 REFERENCE 3 (bases 1 to 10962)  
 AUTHORS Castle,E.  
 JOURNAL Unpublished  
 REFERENCE 4 (bases 67 to 10485)  
 AUTHORS Castle,E., Leidner,U., Nowak,T., Wengler,G. and Wengler,G.  
 TITLE Primary structure of the West Nile flavivirus genome region coding  
 for all nonstructural proteins  
 JOURNAL Virology 149 (1), 10-26 (1986)  
 PUBMED 3753811  
 REFERENCE 5 (bases 1 to 10962)  
 AUTHORS Yamshchikov,V.F., Wengler,G., Perelygin,A.A., Brinton,M.A. and  
 Compans,R.W.  
 TITLE An infectious clone of West Nile flavivirus  
 JOURNAL Virology (2000) In press  
 REFERENCE 6 (bases 1 to 10962)  
 AUTHORS Castle,E.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-AUG-1993) Justus-Liebig-Universitat Giessen, Institut  
 fur Virologie, 35392, Giessen, Germany  
 REFERENCE 7 (bases 1 to 10962)  
 AUTHORS Yamshchikov,V.F.

TITLE Direct Submission  
 JOURNAL Submitted (01-DEC-2000) University of Virginia Health Sciences  
 Centre, Department of Internal Medicine/GI, Charlottesville, VA  
 22906  
 COMMENT On Dec 1, 2000 this sequence version replaced gi:336167.  
 Draft entry and sequence in computer readable form for  
 [1], [2], [4], [3] kindly provided by E.Castle, 12-NOV-1985. The West  
 Nile viral genome consists of a 42S viral RNA. The amino-terminal  
 ends of the structural proteins were experimentally determined. An  
 'atg' codon is located at positions 142-144, which could be used  
 for an alternative initiation of translation for V2. The  
 carboxy-terminal ends of the proteins reported here were not yet  
 precisely defined.  
 FEATURES Location/Qualifiers  
 source 1. .10962  
   /organism="West Nile virus"  
   /virion  
   /mol\_type="genomic RNA"  
   /db\_xref="taxon:11082"  
   /clone="33/G8; 34/F6"  
 CDS 97. .10389  
   /codon\_start=1  
   /product="polyprotein precursor"  
   /protein\_id="AAA48498.2"  
   /db\_xref="GI:11497620"  
   /translation="MSKKPGGPGKNRAVNMLKRGMPRLSLIGLKRAMLSLIDKGKPI  
   RFVLALLAFFFRTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSTKQ  
   KKRGGTAGFTILLGLIACAGAVTLSNFQGVMMTVNATDVTDVITIPTAAGKNLCIVR  
   AMDVGYLCEDTITYECPVLAAGNDPEDIDCWCTKSSVYVRYGRCTKTRHSRRSRSLT  
   VQTHGESTLANKKGAWLDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV  
   VFAILLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLLEGDSCTIMSKDKPTIDVK  
   MMNMEAANLADVRSYCYLASVSDLSTRAACPTMGEAHNEKRADPAFVCKQGVVDRGWG  
   NGCGLFGKGSIDTCAKFACTTKATGWIIQKENIKYEVAIFVHGPTTVESHGKIGATQA  
   GRFSITPSAPSYTTLKGEYGEVTVDCEPRSGIDTSAYVMSVGEKSFLVHREWFMDLN  
   LPWSSAGSTTWRNRETLMEEFEPPHATKQSVALGSQEGALHQALAGAIPEFSSNTVK  
   LTSGHLKCRVKEMLQLKGTTYGVCSKAFFKARTPADTGHGTVVLELQYTGTGPKV  
   PISSVASNLDTPVGRLVTVPFVSATANSKVLIELEPPFGDSYIVVGRGEQQINHH  
   WHKSGSSSIGKAFTTLLRGAQRLAALGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLF  
   GGMSWITQGLLGALLLWMGINARDRSIAMTFLAVGGVLLFLSVNVHADTGCAIDIQRQ  
   ELRCGSGVFIHNDVEAWMDRYKFYPETPQGLAKIIQKAHEGVCGLRSVSRLEHQMWE  
   AIKDELNTLLKENGVDLSVVEKQNGMYKAAPKRLAATTEKLEMGWKAWGKSIIFAPE  
   LANNTFVIDGPETEECPTANRAWNSMEVEDFGFLSTRMFLRIETNTTECDSKIIG  
   TAVKNNMAVHSDLSYWIIESGLNDTWKLERAVLGEVKSCTWPEHTLWGDGVLSDLII  
   PITLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTISDSCEHRGPAARTT  
   TESGKLITDWCCRSCTLPPPLRFQTEENGWYGMEIRPTRHDEKTLVQSRVNAYNADMID  
   PFQLGLMVVFLATQEVLRKRTAKISIPAIMLALLVLVFGGITYTDVLRVILVGAAF  
   AEANSGGDVVHLALMATFKIQPVFLVASFLKARWTNQESILLMLAAFFQMAYYDAKN  
   VLSWEVPDVNLNSVAWMILRAISFTNTSNNVVPLLALLTPGLKCLNLDVYRLLL  
   GVGSLIKEKRSSAAKKGACLICLALASTGVFNPMILAAGLMACDPNRKRGWPATEVM  
   TAVGLMFIAVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDWIERTADITWESDA  
   EITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMRLMACLAI SAYTPWAILPSVIGFWI  
   TLQYTKRGGVWLDTSPKEYKKGDTTGVYRIMTRGLLGSYQAGAGVMVEGVFHTLW  
   TTKGAAALMSGEGRLDPYWGSVKEDRLLCYGGPWKLQHKWNHGDEVQMI  
   VVEPGKVNKVQTKPGVFKTPPEGEIGAVTLDYPTGTSGSPIVDKNGDVIGLYGNGV  
   IMPNGSYISAIVQGERMEEPAPAGFEPEMLRKKQITVLDLHPGAGKTRKILPQI  
   IKEAINKRLRTAVLAPT RVVAAEMSEALRGLPIRYQTSAVREHSGNEIVDVM  
   CHATLTHRLMSPHRVPNYNLFIMDEAHFTDPASIAARGYIATKVELGE  
   AAAIFMTATPPGTSDPFPESNAPISDMQTEIP"

DRAWNTGYEWITEYVGKTVWFVPSVKMGNEIALCLQRAGKKVIQLNRKSYETEYPCK  
NDDWDFITTDISEMGANFKASRVIDSRKSVKPTTIEEGDGRVLGEPSTAASAAQ  
RRGRIGRNPSPQVGDECYGGHTNEEDDSNAHTEARIMLDNINMPNGLVAQLYQPERE  
KVYTMGEYRLRGEERKNFLEFLRTADLPVWLAYKVAAGISYHDRKWCDFGPRTNTI  
LEDNNNEVEVITKLERKILRPRWADARVYSDHQALKSFKDASGKRSQIGLVVLGRM  
PEHFMVKTWEALDTMYVVATAEKGGRAHRMAELPDALQTIVLIALLSVMSLGFFL  
LMQRKGIGKIGLGGVILGAATFFCWMAEVPGTKIAGMLLSLLMIVLIPPEKQRSQ  
TDNQLAVFLICVLTIVGAVAANEMGWLDKTKNDIGSLLGHREARETTLGVESFLDL  
RPATAWSLYAVTTAVLTPLLKHLITSODYINTSLTSINVQASALFTLARGFPFDVGVS  
ALLLAVGCWGQVTLTWTAAALLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNVVVD  
GIVATDVPELERTPVMQKKVGQIIILILVSMAAVVNPSVRTVREAGILTTAAVTLW  
ENGASSVWNATTAIGLCHIMRGGWLSCLSIMWTLIKNEKPGKLKRGGAKGRTLGEVWK  
ERLNHMTKEEFTRYRKEAITEVDRSAAKHARREGNITGGHPVSRGTAKLRWLVERRFL  
EPVGKVVVLGCGRRGCYYMATQKRVQEVKGYTKGGPGHEEPQLVQSYGWNIVTMKSG  
VDVFYRPSEASDTLLCDIGESSSSAEVEEHTVRVLEMVEDWLHRGPKEFCIKVLCPY  
MPKVIKMETLQRYYGGGLIRNPLSRNSTHEMYWVSHASGNIVHSVNMTSQVLLGRME  
KKTWKGPQFEEDVNLGSCTRAGKPLLNSDTSKIKNRIERLKKEYSSTWHQDANHPYR  
TWNYHGSYEVKPTGSASSLVNGVVRLLSKPWDITNVTTMAMTDTTPFGQQRVFKEKV  
DTKAPEPPGVKYVLNETTNWLWAFLARDKKPRMCSSREEFIGKVNSNALGAMFEEQN  
QWKNAREAVEDPKFWEMVDEEREAHLRGECNTCIYNMMGKREKKPGEFGKAKGSRAIW  
FMWLGARFLEFEALGFLNEDHWLGRKNSSGGVEGLGLQKLGYILKEVGTKPGGVYAD  
DTAGWDTRITKADLENEAKVLELLDGEHRRLARSIIELTYRHKKVVKMRPAADGKTV  
DVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLGKGKGPVRTW  
LFENGEERLRSMAVSGDDCVVKPLDDRFATSLHFLNAMS KVRKDIQEWPSTGWDWQ  
QVPFCNSHFTELIMKDGRTLVVPCRGQDELIGRARISPAGGNVRDTACLAKS YAQMW  
LLYFHRRLRLMANAICSAPANWVPTGRTTWSIHAKGEWMTTEDMLAVWNRVWIEE  
NEWMEDKTPVERWSDVPSKGKREDIWCGLIGTRTRATWAENIHVAINVQRSVIGEEK  
YVDYMSLRRYEDTIVVEDTVL"

```
mat_peptide    97. .372
               /product="V2 (14kd core protein)"
sig_peptide   409. .465
               /note="NV2 signal peptide"
mat_peptide   466. .765
               /product="NV2 (20.5 kd membrane-associated glycoprotein)"
mat_peptide   742. .765
               /product="V1 (7 kd membrane-associated nonglycosylated
protein"
sig_peptide   919. .966
               /note="V3 signal peptide"
mat_peptide   967. .2457
               /product="V3 (50 kd membrane-associated glycoprotein;
putative); putative"
sig_peptide   2386. .2457
               /note="nonstructural protein NV4 signal peptide"
mat_peptide   2458. .6426
               /product="nonstructural protein NV4"
mat_peptide   7834. .10380
               /product="nonstructural protein NV5"
```

## ORIGIN

Query Match 86.8%; Score 1319.4; DB 10; Length 10962;  
Best Local Similarity 92.5%; Pred. No. 0;  
Matches 1402; Conservative 0; Mismatches 101; Indels 12; Gaps 1;

```

Qy          1 CGGAATTCTTCAACTGTTAGGAATGAGCAACAGGGACTCCTGGAGGGAGTGCTG 60
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db         956 CAGCATACAGCTTCAACTGTTAGGAATGAGTAACAGAGACTCCTGGAGGGAGTGCTG 1015

```

Qy 61 GAGCTACATGGGTTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAAG 120  
Db 1016 ||||||| 1016 GAGCTACATGGGTTGATCTGGTACTGGAAGGCGATAGTTGTGTGACCATAATGTCAAAAG 1075  
  
Qy 121 ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAATCTCGCAGATGTGC 180  
Db 1076 ||||||| 1076 ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCCACCTCGCAGATGTGC 1135  
  
Qy 181 GTAGCTACTGCTACTTAGCTCGGTCACTGATCTGTCAACAAAAGCCCGTGTCCAACCA 240  
Db 1136 ||||||| 1136 GCAGTTACTGTTACCTAGCTCGGTCACTGACTTGTCAACAAGAGCTCGTGTCCAACCA 1195  
  
Qy 241 TGGGTGAAGCTCACAACGAGAAAAGAGCCGACCCCTGCCTTGTTGCAAGCAAGGCGTCG 300  
Db 1196 ||||||| 1196 TGGGTGAAGCCCACAAACGAGAAAAGAGCTGACCCCCGCTTCGTTGCAAGCAAGGCGTTG 1255  
  
Qy 301 TAGACAGAGGATGGGAAATGGATGCGGACTGTTGGAAAGGGGAGCATTGACACATGTG 360  
Db 1256 ||||||| 1256 TGGACAGAGGATGGGAAATGGCTGCGGACTGTTGGAAAGGGGAGCATTGACACATGTG 1315  
  
Qy 361 CAAAGTTGCCTGTACAACCAAGGCAACTGGTTGATTATCCAGAAGGAAAACATCAAGT 420  
Db 1316 ||||||| 1316 CGAAGTTGCCTGTACAACCAAAGCAACTGGATGGATCATCCAGAAGGAAAACATCAAGT 1375  
  
Qy 421 ACGAGGTTGCCATATTGTGCATGGCCGACGACTGTCGAATCACATGGCAATTATTCAA 480  
Db 1376 ||||||| 1376 ATGAGGTTGCCATATTGTGCATGGCCGACGACCGTTGAATCTCATGGCA----- 1426  
  
Qy 481 CACAGATAGGGCTACCCAAAGCAGGAAGGTTCAGCATAACTCCATGGCACCATCCTACA 540  
Db 1427 ||||||| 1427 ---AGATAGGGGCCACCCAGGCTGGAAGATTCACTTACACTCCATGGCACCATCTTACA 1483  
  
Qy 541 CGCTGAAGTTGGGTGAGTATGGTGAGGTACAGTTGACTGTGAGCCACGGTCAGGAATAG 600  
Db 1484 ||||||| 1484 CGCTAAAGTTGGGTGAGTATGGTGAGGTACGGTTACGGTTGATTGTGAGCCACGGTCAGGAATAG 1543  
  
Qy 601 ACACTAGCGCTTACTACGTTATGTCAGTGGTGCGAACGTCCTTCTGGTCACCGAGAAT 660  
Db 1544 ||||||| 1544 ACACCAGCGCCTATTACGTTATGTCAGTGGTGAGAAGTCCTCCTGGTCACCGAGAAT 1603  
  
Qy 661 GGTTTATGGACCTGAACCTTCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG 720  
Db 1604 ||||||| 1604 GGTTTATGGATCTGAACCTGCCATGGAGCAGTGCTGGAAGCACCACGTGGAGGAACCGGG 1663  
  
Qy 721 AAACACTGATGGAGTTGAAGAACCTCATGCCACCAAACAATCTGTCGTAGCTCTAGGGT 780  
Db 1664 ||||||| 1664 AAACACTGATGGAGTTGAAGAACCTCATGCCACCAAACAATCTGTTGTGGCTCTAGGGT 1723  
  
Qy 781 CGCAGGAAGGTGCCCTGCACCAAGCTCTGGCTGGAGCAATTCTGTTGAGTTCTCAAGCA 840  
Db 1724 ||||||| 1724 CGCAGGAAGGTGCCCTGCACCAAGCTCTGGCCGGAGCGATTCTGTTGAGTTCTCAAGCA 1783  
  
Qy 841 ACACGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC 900  
Db 1784 ||||||| 1784 ACACGTGAAGTTGACATCAGGACATCTGAAGTGTGGGTGAAGATGGAGAAGTTGCAGC 1843

Qy 901 TGAAGGAAACAACATATGGTGTATGCTCAAAGCATTCAAATTGCTAGGACTCCCGCTG 960  
 ||||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1844 TGAAGGAAACAACATATGGAGTATGTTCAAAGCGTTCAAATTGCTAGGACTCCCGCTG 1903  
  
 Qy 961 ACACGGTCATGGAACGGTGGTGGAACTGCAGTATACCGGAAAAGACGGGCCTTGCA 1020  
 ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||| |||||  
 Db 1904 ACACGGCCACGGAACGGTGGTGGAACTGCAATATACCGAACAGACGGTCCCTGCA 1963  
  
 Qy 1021 AAGTGCCCATTCTCTGTGGCTCCCTGAACGACCTACACCCGTTGGAAGGCTGGTGA 1080  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1964 AAGTGCCCATTCTCCGTAGCTCCCTGAATGACCTCACACCTGTTGGAAGACTGGTGA 2023  
  
 Qy 1081 CTGTGAATCCATTGTGTCTGTGGCTACGCCAAGTCAAGGTTTGATTGAACCTGAAC 1140  
 | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2024 CCGTGAATCCATTGTGTCTGTGGCCACAGCCAAGTCAAGGTTTGATTGAACCTGAAC 2083  
  
 Qy 1141 CCCCGTTAGTGAACCTTACATCGTGGTGGGAGAGGAGAACAGCAGATAAACCAACT 1200  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2084 CCCCGTTGGTGAACCTTACATCGTGGTGGAGAGGAGAACAGCAGATAAACCATCACT 2143  
  
 Qy 1201 GGCACAAATCTGGAGCAGTATTGAAAGGCTTCACCACTACACTCAGAGGAGCTAAC 1260  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2144 GGCACAAATCTGGAGCAGCATTGAAAGGCCCTTACCAACACTCAGAGGAGCTAAC 2203  
  
 Qy 1261 GACTTGCAGCTCTGGAGACACTGCCTGGGATTTGGATCAGTCGGAGGGGTTTCACCT 1320  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2204 GACTCGCAGCTCTGGAGATACTGCTTGGGATTTGGATCAGTTGGAGGGGTTTCACCT 2263  
  
 Qy 1321 CGGTAGGGAAAGCCATACACCAAGTTTGAGGAGCCTTAGATCACTCTTGGAGGGA 1380  
 | ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2264 CAGTGGGGAAAGCCATACACCAAGTCTTGGAGGAGCTTAGATCACTCTTGGAGGGA 2323  
  
 Qy 1381 TGTCTGGATCACACAGGGCTTCTGGAGCTCTGCTGTGGATGGAATTAACGCC 1440  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2324 TGTCTGGATCACACAGGGACTTCTGGAGCTCTGTTGTGGATGGAATCAATGCC 2383  
  
 Qy 1441 GTGACAGGTCAATTGCTATGACGTTCTTGCCTGGAGGAGTCTGCTCTCCTTCGG 1500  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2384 GTGACAGGTCAATTGCTATGACGTTCTTGCCTGGAGGAGTTTGCTCTCCTTCGG 2443  
  
 Qy 1501 TCAACGTCCATGCTG 1515  
 ||||| |||||  
 Db 2444 TCAACGTCCATGCTG 2458

#### RESULT 7

AR365300

LOCUS AR365300 1491 bp DNA linear PAT 03-SEP-2003  
 DEFINITION Sequence 3 from patent US 5486473.  
 ACCESSION AR365300  
 VERSION AR365300.1 GI:34428831  
 KEYWORDS .  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 Unclassified.  
 REFERENCE 1 (bases 1 to 1491)  
 AUTHORS Fujita,H., Yoshida,I., Takagi,M., Manabe,S. and Fukai,K.

TITLE A DNA coding for a Flavivirus antigen  
JOURNAL Patent: US 5486473-A 3 23-JAN-1996;  
The Research Foundation for Microbial Diseases of Osaka University;  
Osaka;  
JPX;

FEATURES Location/Qualifiers  
source 1. .1491  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Query Match 86.3%; Score 1312.2; DB 2; Length 1491;  
Best Local Similarity 92.7%; Pred. No. 0;  
Matches 1393; Conservative 0; Mismatches 98; Indels 12; Gaps 1;

Qy	12	TTCAACTGTTAGGAATGAGCAACAGGGACTCCTGGAGGGAGTGTCTGGAGCTACATGG	71
Db	1	TTCAACTGTTAGGAATGAGTAACAGAGACTCCTGGAGGGAGTGTCTGGAGCTACATGG	60
Qy	72	GTTGATCTGGTACTGGAAGGGAGACAGTTGTGTGACCATAATGTCAAAAGACAAGCCAACC	131
Db	61	GTTGATCTGGTACTGGAAGGCATAGTTGTGTGACCATAATGTCAAAAGACAAGCCAACC	120
Qy	132	ATTGATGTCAAAATGATGAACATGGAAGCAGCTAATCTCGCAGATGTGCGTAGCTACTGC	191
Db	121	ATTGATGTCAAAATGATGAACATGGAAGCAGCCAACCTCGCAGATGTGCGCAGTTACTGT	180
Qy	192	TACTTAGCTCGGTCACTGATCTGTCAACAAAAGCCCGTGTCCAACCATGGTGAAGCT	251
Db	181	TACCTAGCTCGGTCACTGACTTGTCAACAAGAGCTGCGTGTCCAACCATGGTGAAGCC	240
Qy	252	CACAACGAGAAAAGAGCCGACCCCTGCCTTGTGCAAGCAAGGCCGTAGACAGAGGA	311
Db	241	CACAACGAGAAAAGAGCTGACCCCGCTTCGTTGCAAGCAAGGCCGTGTGGACAGAGGA	300
Qy	312	TGGGGAAATGGATGCGGACTGTTGGAAAGGGGAGCATTGACACATGTGCAAAGTTGCC	371
Db	301	TGGGGAAATGGCTGCGGACTGTTGGAAAGGGGAGCATTGACACATGTGCGAAGTTGCC	360
Qy	372	TGTACAACCAAGGCAACTGGTTGGATTATCCAGAAGGAAAACATCAAGTACGAGGTTGCC	431
Db	361	TGTACAACCAAGCAACTGGATGGATCATCCAGAAGGAAAACATCAAGTATGAGGTTGCC	420
Qy	432	ATATTTGTGCATGGCCGACGACTGTCGAATCACATGCCATTATTCAACACAGATAGGG	491
Db	421	ATATTTGTGCATGGCCGACGACCGTTGAATCTCATGGCA-----AGATAGGG	468
Qy	492	GCTACCCAAGCAGGAAGGTTCAGCATAACTCCATCGGCACCATCCTACACGCTGAAGTTG	551
Db	469	GCCACCCAGGCTGGAAGATTCACTGATAACTCCATCGGCACCATCTTACACGCTAAAGTTG	528
Qy	552	GGTGAGTATGGTGGAGGTTACGGTTGATTGTGAGCCACGGTCAGGAATAGACACTAGCGCT	611
Db	529	GGTGAGTATGGTGGAGGTTACGGTTGATTGTGAGCCACGGTCAGGAATAGACACCAGCGCC	588
Qy	612	TACTACGTTATGTCAGTGGGTGCGAAGTCCTCTGGTTACCGAGAATGGTTATGGAC	671

Db 589 TATTACGTTATGTCAGTTGGTGAGAAGTCCTCCTGGTCACCGAGAATGGTTATGGAT 648  
Qy 672 CTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACTGGAGGAACCGGGAAACACTGATG 731  
|||  
Db 649 CTGAACCTGCCATGGAGCAGTGCTGGAAGCACCACGTGGAGGAACCGGGAAACACTGATG 708  
Qy 732 GAGTTGAAGAACCTCATGCCACCAAACAATCTGTCGTAGCTCTAGGTCGCAGGAAGGT 791  
|||  
Db 709 GAGTTGAAGAACCTCATGCCACCAAACAATCTGTTGTGGCTCTAGGTCGCAGGAAGGT 768  
Qy 792 GCCTTGCACCAAGCTCTGGCTGGAGCAATTCTGTTGAGTTCTCAAGCAACACTGTGAAG 851  
|||  
Db 769 GCGTTGCACCAAGCTCTGGCCGGAGCGATTCTGTTGAGTTCTCAAGCAACACTGTGAAG 828  
Qy 852 TTGACATCAGGACATCTGAAGTGTAGGGTAAGATGGAGAAGTTGCAGCTGAAGGGAAACA 911  
|||  
Db 829 TTGACATCAGGACATCTGAAGTGTGGGTGAAGATGGAGAAGTTGCAGCTGAAGGGAAACA 888  
Qy 912 ACATATGGTGTATGCTAAAAGCATTCAAATTGCTAGGACTCCGCTGACACTGGTCAT 971  
|||  
Db 889 ACATATGGAGTATGTTAAAAGCATTCAAATTGCTAGGACTCCGCTGACACTGGCAC 948  
Qy 972 GGAACGGTGGTGTGGAACTGCAGTATAACCGAAAAGACGGGCCTGCAAAGTGCCATT 1031  
|||  
Db 949 GGAACGGTGGTGTGGAACTGCAATATAACCGGAACAGACGGTCCCTGCAAAGTGCCATT 1008  
Qy 1032 TCTTCTGTGGCTTCCCTGAACGACCTTACACCCGTTGAAAGGCTGGTACTGTGAATCCA 1091  
|||  
Db 1009 TCTTCCGTAGCTTCCCTGAATGACCTCACACCTGTTGAAAGACTGGTACCGTGAATCCA 1068  
Qy 1092 TTTGTGTCTGTGGCTACGGCAACTCGAAGGTTTGATTGAACTCGAACCCCCGTTAGT 1151  
|||  
Db 1069 TTTGTGTCTGTGGCCACAGCCAACCTCGAAGGTTTGATTGAACTCGAACCCCCGTTGGT 1128  
Qy 1152 GACTCTTACATCGTGGGGAGAGGAGAACAGCAGATAAACCAACTGGCACAAATCT 1211  
|||  
Db 1129 GACTCTTACATCGTGGGGAGAGGAGAACAGCAGATAAACCATCACTGGCACAAATCT 1188  
Qy 1212 GGGAGCAGTATTGAAAGGCTTCACCAACTACACTCAGAGGAGCTAACGACTTGCAGCT 1271  
|||  
Db 1189 GGGAGCAGCATTGAAAGGCCTTACCAACACTCAGAGGAGCTAACGACTCGCAGCT 1248  
Qy 1272 CTTGGAGACACTGCCTGGATTGGATCAGTCGGAGGGTTTCACCTCGTAGGGAAA 1331  
|||  
Db 1249 CTTGGAGATACTGCTTGGATTGGATCAGTTGGAGGGTTTCACCTCAGTGGGGAAA 1308  
Qy 1332 GCCATACACCAAGTTTGGAGGAGCCTTAGATCACTCTTGGAGGGATGTCCTGGATC 1391  
|||  
Db 1309 GCCATACACCAAGTCTTGGAGGAGCTTTAGATCACTCTTGGAGGGATGTCCTGGATC 1368  
Qy 1392 ACACAGGGCTCTGGAGCTCTGCTGTGGATGGAATTACGCCGTGACAGGTCA 1451  
|||  
Db 1369 ACACAGGGACTCTGGAGCTCTGTTGGAGGAGTTAGATCAATGCCGTGACAGGTCA 1428  
Qy 1452 ATTGCTATGACGTTCTTGCAGGTGGAGGAGCTTGCTCTGGTCAACGTCCAT 1511  
|||  
Db 1429 ATTGCTATGACGTTCTTGCAGGTGGAGGAGTTGCTCTTGCAGGTCAACGTCCAT 1488

Qy 1512 GCT 1514  
|||  
Db 1489 GCT 1491

RESULT 8  
AF394221  
LOCUS AF394221 1430 bp mRNA linear VRL 03-MAY-2002  
DEFINITION West Nile virus isolate B956 polyprotein mRNA, envelope glycoprotein E and nonstructural protein 1 region, partial cds.  
ACCESSION AF394221  
VERSION AF394221.1 GI:20428494  
KEYWORDS .  
SOURCE West Nile virus  
ORGANISM West Nile virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.  
REFERENCE 1 (bases 1 to 1430)  
AUTHORS Briese,T., Rambaut,A., Pathmajeyan,M., Bishara,J., Weinberger,M., Pitlik,S. and Lipkin,W.I.  
TITLE Phylogenetic analysis of a human isolate from the 2000 Israel West Nile virus epidemic  
JOURNAL Emerging Infect. Dis. 8 (5), 528-531 (2002)  
PUBMED 11996693  
REFERENCE 2 (bases 1 to 1430)  
AUTHORS Briese,T., Jordan,I., Pathmajeyan,M. and Lipkin,W.I.  
TITLE Direct Submission  
JOURNAL Submitted (21-JUN-2001) Emerging Diseases Laboratory, Microbiology & Molecular Genetics, and Neurology, University California Irvine, 3107 Gillespie Neuroscience Building, Irvine, CA 92697-4292, USA  
FEATURES Location/Qualifiers  
source 1. .1430  
/organism="West Nile virus"  
/mol\_type="mRNA"  
/isolate="B956"  
/db\_xref="taxon:11082"  
/country="Uganda"  
/note="isolated from human serum in 1937; kindly provided by Bob Tesh, University of Texas Medical Br., Galveston, TX, USA"  
CDS <1. .>1430  
/codon\_start=3  
/product="polyprotein"  
/protein\_id="AAM21944.1"  
/db\_xref="GI:20428495"  
/translation="KRADPAFVKQGVVDRGWGNGCGLFGKGSIDTCAKFACTTKATG  
WIQKENIKYEVAIFVHGPTTVEHKGIGATQAGRFSITPSAPSYTTLKLGEYGEVTVD  
CEPRSGIDTSAYYVMSVGAKSFLVHREWFMDLNLPWSAGSTTWRNRETLVEFEELPH  
TKQSVALGSQEGALHQALAGAIPEFSSNTVKLTSGHLKCRVKMEKLQLKGTTYGVC  
SKAFKFARTPADTGHGTVVLELQYTGTDPCKVPISSVASLNDLTPVGRILTVNPFVS  
VATANSKVLLIELEPPFGDSYIVVGRGEQQINHHWHKGSSSIGKAFTTLRGAQRLLAAL  
GDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDR  
SIAMTFLAVGGVLLFLSVNVHADTGCAIDIGRQELRCGSVFIHNDVEAWMDRYKFYP  
ETPQGLAKIQKAHAEGVCGLRSVR"  
mat\_peptide <1. .1244  
/product="envelope glycoprotein E"

mat\_peptide 1245. .>1430  
/product="nonstructural protein 1"  
/note="NS1"

ORIGIN

Query Match 71.9%; Score 1093.4; DB 10; Length 1430;  
Best Local Similarity 92.6%; Pred. No. 0;  
Matches 1164; Conservative 0; Mismatches 81; Indels 12; Gaps 1;

Qy 259 AGAAAAGAGCCGACCCCTGCCTTGGCAAGCAAGGCCTCGTAGACAGAGGATGGGGGA 318  
Db 1 AGAAAAGAGCTGACCCCGCCTCGTTGCAAGCAAGGCCTGTGGACAGAGGATGGGGAA 60

Qy 319 ATGGATGCCGACTGTTGGAAAGGGGAGCATTGACACATGTGCAAAGTTGCCTGTACAA 378  
Db 61 ATGGCTGCCGACTGTTGGAAAGGGGAGCATTGACACATGTGCGAAGTTGCCTGTACAA 120

Qy 379 CCAAGGCAACTGGTTGGATTATCAGAAGGAAACATCAAGTACGAGGTTGCCATATTG 438  
Db 121 CCAAAGCAACTGGATGGATCATCCAGAAGGAAACATCAAGTATGAGGTTGCCATATTG 180

Qy 439 TGCATGGCCCGACGACTGTCGAATCACATGGCAATTATTCAACACAGATAGGGCTACCC 498  
Db 181 TGCATGGCCCGACGACCCTGAATCTCATGGCA-----AGATAGGGCCACCC 228

Qy 499 AAGCAGGAAGGTTCAGCATAACTCCATCGGCACCATCCTACACGCTGAAGTTGGTGAGT 558  
Db 229 AGGCTGGAAGATTCACTTAACTCCATCGGCACCATCTTACACGCTAAAGTTGGTGAGT 288

Qy 559 ATGGTGAGGTACAGTTGACTGTGAGCCACGGTCAGGAATAGACACTAGCGCTTACTACG 618  
Db 289 ATGGTGAGGTACGGTTGATTGTGAGCCACGGTCAGGAATAGACACTAGCGCTTACTACG 348

Qy 619 TTATGTCAGTGGTGCAGTCCTTCTGGTCACCGAGAATGGTTATGGACCTGAACC 678  
Db 349 TTATGTCAGTGGTGCAGTCCTTCTGGTCACCGAGAATGGTTATGGACCTGAACC 408

Qy 679 TTCCATGGAGTAGCGCTGGAAAGCACAACGTGGAGGAACCGGGAAACACTGATGGAGTTG 738  
Db 409 TGCCATGGAGCAGTGTGGAAAGCACCACGTGGAGGAACCGGGAAACACTGGTGGAGTTG 468

Qy 739 AAGAACCTCATGCCACCAAAACATCTGTCAGCTAGCTCTAGGGTCGCAGGAAGGTGCCTTGC 798  
Db 469 AAGAACCTCATGCCACCAAAACATCTGTTGTGGCTCTAGGGTCGCAGGAAGGTGCCTTGC 528

Qy 799 ACCAAGCTCTGGCTGGAGCAATTCTGTTGAGTTCTCAAGCAACACTGTGAAGTTGACAT 858  
Db 529 ACCAAGCTCTGGCGGAGCATTCTGTTGAGTTCTCAAGCAACACTGTGAAGTTGACAT 588

Qy 859 CAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGCTGAAGGAAACAACATATG 918  
Db 589 CAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGCTGAAGGAAACAACATATG 648

Qy 919 GTGTATGCTAAAAGCATTCAAATTGCTAGGACTCCGCTGACACTGGTCAAGGAAACGG 978  
Db 649 GAGTATGTTCAAAAGCGTTCAAATTGCTAGGACTCCGCTGACACTGGCACGGAACGG 708

Qy	979	TGGTGCTGGAACGTGAGTATACCGGAAAAGACGGGCCTGCAAAGTGCCATTCTCTG	1038
Db	709	TGGTGTGGAACGTGCAATATACCGAACAGACGGCCTGCAAAGTGCCATTCTCCG	768
Qy	1039	TGGCTCCCTGAACGACCTTACACCGTTGAAGGCTGGTACTGTGAATCCATTGTGT	1098
Db	769	TAGCTCCCTGAATGACCTCACACCTGTTGAAGACTGGTACCGTGAATCATTGTGT	828
Qy	1099	CTGTGGCTACGGCCAACTCGAAGGTTTGATTGAACTCGAACCCCCGTTAGTGACTCTT	1158
Db	829	CTGTGGCCACAGCCAACTCGAAGGTTTGATTGAACTCGAACCCCCGTTGGTACTCTT	888
Qy	1159	ACATCGTGGTGGGAGAGGAGAACAGCAGATAAACCACTGGCACAAATCTGGGAGCA	1218
Db	889	ACATCGTGGTGGGAGAGGAGAACAGCAGATAAACCATCACTGGCACAAATCTGGGAGCA	948
Qy	1219	GTATTGAAAGGCTTCACCACTACACTCAGAGGAGCTAACGACTTGCAGCTCTGGAG	1278
Db	949	GCATTGAAAGGCCTTACCAACCACACTCAGAGGAGCTAACGACTCGCAGCTCTGGAG	1008
Qy	1279	ACACTGCCTGGGATTTGGATCAGTCGGAGGGTTTCACCTCGTAGGGAAAGCCATAC	1338
Db	1009	ATACTGCTTGGGATTTGGATCAGTTGGAGGGTTTCACCTCAGTGGGAAAGCCATAC	1068
Qy	1339	ACCAAGTTTGGAGGAGCCTTAGATCACTCTTGGAGGGATGCTGGATCACACAGG	1398
Db	1069	ACCAAGTCTTGGAGGAGCTTAGATCACTCTTGGAGGGATGCTGGATCACACAGG	1128
Qy	1399	GGCTTCTGGGAGCTCTGCTGTGGATGGAATTAAACGCCGTGACAGGTCAATTGCTA	1458
Db	1129	GACTTCTGGGAGCTCTGTTGGATGGAATTAAACGCCGTGACAGGTCAATTGCTA	1188
Qy	1459	TGACGTTCTTGCAGGTTGGAGGAGCTTGCTCTTCCTTCGGTCAACGTCCATGCTG	1515
Db	1189	TGACGTTCTTGCAGGTTGGAGGAGTTTGCTCTTCCTTCGGTCAACGTCCATGCTG	1245

RESULT 9

AY701413

LOCUS AY701413 10945 bp RNA linear VRL 08-FEB-2005  
 DEFINITION West Nile virus strain 04.05 polyprotein gene, complete cds.  
 ACCESSION AY701413  
 VERSION AY701413.1 GI:51011375  
 KEYWORDS .  
 SOURCE West Nile virus (WNV)  
 ORGANISM West Nile virus  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 Flavivirus; Japanese encephalitis virus group.  
 REFERENCE 1 (bases 1 to 10945)  
 AUTHORS Schuffenecker,I., Peyrefitte,C.N., el Harrak,M., Murri,S., Leblond,A. and Zeller,H.G.  
 TITLE West Nile virus in Morocco, 2003  
 JOURNAL Emerging Infect. Dis. 11 (2), 306-309 (2005)  
 PUBMED 15752452  
 REFERENCE 2 (bases 1 to 10945)  
 AUTHORS Schuffenecker,I., Murri,S. and Zeller,H.G.  
 TITLE Direct Submission

JOURNAL Submitted (29-JUL-2004) CNR Arbovirus, Institut Pasteur, 21 Avenue Tony Garnier, Lyon cedex 07 69365, France

FEATURES Location/Qualifiers

source 1. .10945  
/organism="West Nile virus"  
/mol\_type="genomic RNA"  
/strain="04.05"  
/isolation\_source="brain of horse with encephalitis"  
/specific\_host="horse"  
/db\_xref="taxon:11082"  
/country="Morocco"  
/collection\_date="2003"

CDS 55. .10356  
/codon\_start=1  
/product="polyprotein"  
/protein\_id="AAT92099.1"  
/db\_xref="GI:51011376"  
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDKGKGPI  
RFVLALLAFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ  
KKRGGKTGIAVMIGLIASVGAVTLSNFQGVMMTVNATDVTDVITIPTAACGNLCIVR  
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAYVYRYGRCTKTRHSRSRRSLT  
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV  
VFVVLVLLVAPAYSFNCLGMSNRDFLEGVSGATWDLVLEGDSCVTIMSKDKPTIDVK  
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRCQGVVDRGWG  
NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVEHGNYSTQIG  
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYMTVGTKTFLVHREW  
MDLNLPWSSAGSTVWRNRETLMEFEPPHATQSVIALGSQEGALHQALAGAIPEFSS  
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCASKFKLGPADTGHTVVLELQYTGTDG  
PCKVPISSVASLNDLTPVGRLTVPFVSVATANAKVIELEPPFGDSYIVVGRGEQQ  
INHHWHKSGSIGKAFTTLKGQAQRILAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF  
RSLFGGMSWIQGLLGALLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTCAID  
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIQKAHKEGVCGLRSVSRL  
QMWEVKDELNTLLKENGVDLSVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL  
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECD  
KIIGTAIKNNLAIHSDSLWYIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES  
DLIIPVTLAGPRSNNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA  
TRTTESGKLITDWCCRSCTLPPRLRYQTDSCGCWYGMERPQRHDEKTLVQSQVNAYNA  
DMIDPFQLGLLVVFLATQEVLRKWTAKISMPAILIALLVLVFGGITYTDVLRVILV  
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAFFQOMAYY  
DARQILLWEIPDVNLNSLAVAAMILRAITFTTSNVVVPALLPGLRCLNLDVYRIL  
LLMVGIGSLIREKRSAAKKKGASLLCLALASTGLFNPMLAAGLIACDPNRKRGWPA  
TEVMTAVGLMFAIVGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDWMIERTADIS  
ESDAEITGSSEVDVRLLDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVV  
GFWITLQYTKRGGVWLWTPSPKEYKKGDTTGTYRIMTRGLLGSYQAGAGVMVEGV  
TLWHTTKGAALMSGEGRRLDPYWGGSVKEDRLCYGGPWKLQHKWNGQDEVQMV  
VKNVQTKPGVFKTPGEIGAVTLDFTGTSGPSIVDKNGDVIGLYGNGVIMPNGSYIS  
AIVQGERMDEPIPAGEPEPEMLRKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV  
LAPTRVVAEAEALRGLPIRYQTSAVTREHNGNEIVDMCHATLTHRLMSPHRP  
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAI FMTATPPGTSDPFPESN  
SPISDLQ TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKVV  
QLNRSYETEY PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTI  
ITEGEGRVILGEPEAVTAA SAAQRRGRIGRNP  
S QVGDEYCYGGHTNEDDSNFAHWT  
TEARIMLDNINMPNGLIAQFYQ PEREKVYTM  
DGEYRLGEERKNF  
LELLRTADLPV  
WLA  
YK  
VAAAG  
VSYH  
DRRC  
FDGPR  
TNTILED  
NNEVE  
VITKL  
GERK  
ILRPR  
WIDAR  
VYSDH  
QALKAF  
KDFASG  
KRSQIG  
LIEV  
LGK  
MPEH  
FMGKT  
WEAL  
DTM  
YV  
VATA  
EKG  
GRR  
AHR  
MALE  
ELPD  
ALQT  
IALIA  
LLS  
VMT  
MG  
VF  
FLL  
MQR  
KIG  
KIG  
LGG  
VVL  
GVAT  
FFC  
WMAEV  
PGTK  
IAG  
M  
L  
L  
M  
V  
I  
L  
P  
E  
P  
E  
K  
L  
V  
E  
K  
N  
F  
S  
M  
G  
E  
F  
L

LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFDV  
GVSALLLAAGCWGQVTLTVTVAATLLFCHYAYMVGWQAEAMRSAQRRTAACIMKNA  
VVDGIVATDVEPELERTPIMQKKVQIMLILVSLAAVVNPSPVKTVREAGILITA  
TLWENGASSVWNATTAILCHIMRGGWLSCLSITWLTIKNMDKPLKRGGA  
KRTLGEVWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPV  
SRGTAKLRWLVERRFLEPVGKVIDLGCGRGGWCYYMATQKRVQEV  
RGYTKGPGHEEPQLVQSYGWNIVTMKSGVDVFYRPSECCDTLLCDIG  
ESSSSAEEHHRTIRVLEMVEDWLHRCGPREFCVKVLCPYMPKVIE  
EKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRA  
SGNVVHSVNMTSQVLLGRMEKRTWKGPQYE  
EDVNLGSCTRAVGKPLNSDTSKIKNRIERLREYS  
STWHHDENHPYRTWNYHGSYDV  
KPTGSASSLVNGVVRLLSKPWD  
TINVTTMAMTD  
TTPFGQQRVFKEKVDTKAPEP  
PEGVKYVLNETTNWLWAFLAREKRPRMC  
SREEFIRKVN  
SNAALGAMFE  
EQNQWRSARE  
AVEDPKF  
WEMVDE  
EREAHLR  
GECHTCI  
YNMMGKRE  
KKPG  
EGFG  
KAKG  
SR  
AIWF  
MWLGAR  
FLE  
FA  
LGF  
LN  
EDH  
WLGR  
KNSGG  
V  
EG  
GL  
LQ  
KL  
GY  
IL  
REV  
GTR  
PG  
GKI  
YAD  
DTAG  
WDTR  
ITRAD  
LE  
NEAK  
V  
LE  
LD  
GE  
HRR  
LAR  
AI  
I  
ELTY  
RH  
KV  
V  
K  
VMR  
PA  
ADGR  
TVMD  
VIS  
RED  
QR  
GS  
QV  
VT  
Y  
A  
L  
N  
T  
FT  
N  
L  
A  
V  
Q  
L  
V  
R  
M  
E  
G  
E  
G  
V  
I  
G  
P  
D  
D  
V  
E  
K  
L  
T  
K  
G  
G  
P  
K  
V  
RT  
WL  
F  
E  
N  
G  
E  
E  
R  
L  
S  
R  
M  
A  
V  
G  
G  
D  
C  
V  
V  
K  
P  
L  
D  
D  
R  
F  
A  
T  
S  
L  
H  
F  
L  
N  
A  
M  
S  
K  
V  
R  
K  
D  
I  
Q  
E  
W  
K  
P  
S  
T  
G  
W  
Y  
DW  
Q  
Q  
V  
P  
F  
C  
S  
N  
H  
F  
T  
E  
L  
I  
M  
K  
D  
G  
R  
T  
L  
V  
V  
P  
C  
R  
G  
Q  
D  
E  
L  
V  
G  
R  
A  
R  
I  
S  
P  
G  
A  
G  
W  
N  
V  
R  
D  
T  
A  
C  
L  
A  
K  
S  
Y  
A  
Q  
M  
W  
L  
L  
Y  
F  
H  
R  
R  
D  
L  
R  
L  
M  
A  
N  
A  
I  
C  
S  
A  
V  
P  
V  
N  
W  
V  
P  
T  
G  
R  
T  
T  
W  
S  
I  
H  
A  
G  
G  
E  
W  
M  
T  
T  
E  
D  
M  
L  
E  
V  
W  
N  
R  
V  
W  
I  
E  
E  
N  
W  
E  
M  
D  
K  
T  
P  
V  
E  
K  
W  
S  
D  
V  
P  
Y  
S  
G  
K  
R  
E  
D  
I  
W  
C  
G  
S  
L  
I  
G  
T  
R  
A  
R  
A  
T  
W  
A  
E  
N  
I  
Q  
V  
A  
I  
N  
Q  
V  
R  
A  
I  
I  
G  
D  
E  
K  
Y  
V  
D  
Y  
M  
S  
S  
L  
K  
R  
Y  
E  
D  
T  
T  
L  
V  
E  
D  
T  
V  
L

## ORIGIN

Query Match 66.0%; Score 1003; DB 10; Length 10945;  
 Best Local Similarity 78.9%; Pred. No. 0;  
 Matches 1195; Conservative 0; Mismatches 320; Indels 0; Gaps 0;

```

  1 CGGAATTCAAGCTTCAACTGTTAGGAATGAGCAACAGGGACTTCTGGAGGGAGTGTCTG 60
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  914 CAGCCTACAGCTTCAACTGCCTTGGAAATGAGCAACAGAGACTTCTGGAGGGAGTATCTG 973

  61 GAGCTACATGGGTTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAG 120
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  974 GAGCAACATGGGTGGATTTGGTCTCGAAGGCGACAGCTGCGTGAECTATCATGTCCAAGG 1033

  121 ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAACATCGCAGATGTGC 180
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  1034 ACAAGCCCACCATTGATGTGAAGATGATGAATATGGAGGCTGCCAACCTGGCAGAGGTCC 1093

  181 GTAGCTACTGCTACTTAGCTTCGGTCAGTGATCTGTCAACAAAAGCCCGTGTCCAACCA 240
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  1094 GCAGTTATTGCTATTGGCTACCGTCAGCGATCTCTCACCAAGCTGCATGCCGACCA 1153

  241 TGGGTGAAGCTCACAAACGAGAAAAGAGCCGACCCCTGCCTTGTTGCAAGCAAGGCGTCG 300
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  1154 TGGGAGAAGCTCACAAACGACAAACGTGCTGACCCGGCTTGTGTGAGACAAGGAGTGG 1213

  301 TAGACAGAGGATGGGGAATGGATGCGGACTGTTGGAAAGGGGAGCATTGACACATGTG 360
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  1214 TGGACAGAGGCTGGGCAACGGCTGCGGACTATTGGCAAAGGAAGCATTGACACATGCG 1273

  361 CAAAGTTGCCTGTACAACCAAGGCAACTGGTGGATTATCCAGAAGGAAAACATCAAGT 420
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  1274 CCAAATTGCTGCCACCAAGGCAACAGGAAGAACCATCTGAAAGAGAACATCAAGT 1333

  421 ACGAGGTTGCCATATTGTGCATGGCCGACGACTGTCGAATCACATGGCAATTATTCAA 480
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  1334 ATGAAGTGGCCATCTTGTCCATGGACCAACCACTGTGGAGTCGCATGGAAACTACTCCA 1393

```

Qy	481	CACAGATAGGGCTACCCAAGCAGGAAGGTTCAGGCATAACTCCATCGGCCACCCTTACA	540
Db	1394	CACAGATTGGGCCACTCAGGCAGGGAGATTCAAGCATCACTCCTGCAGGCGCTTCATACA	1453
Qy	541	CGCTGAAGTTGGTGAGTATGGTGAGGTACAGTTGACTGTGAGCCACGGTCAGGAATAG	600
Db	1454	CACTAAAGCTGGAGAATATGGAGAAGTGACAGTGGACTGTGAACCACGGTCAGGGATTG	1513
Qy	601	ACACTAGCGCTTACTACGTTATGTCAGTGGTGCGAACGTCTTCTGGTCACCGAGAAT	660
Db	1514	ACACCAATGCCTACTACGTGATGACTGTTGGAACAAAGACGTTTGGTCATCGTGAGT	1573
Qy	661	GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG	720
Db	1574	GGTTCATGGACCTCAACCTCCCTGGAGCAGTGCTGGAAGTACTGTGGAGGAATAGAG	1633
Qy	721	AAACACTGATGGAGTTGAAGAACCTCATGCCACCAAACAATCTGTCGTAGCTCTAGGGT	780
Db	1634	AGACGTTAATGGAGTTGAGGAACCACAGCCACAAAGCAGTCTGTGATAGCACTGGGCT	1693
Qy	781	CGCAGGAAGGTGCCCTGCACCAAGCTCTGGCTGGAGCAATTCTGTTGAGTCTCAAGCA	840
Db	1694	CACAAGAGGGAGCTCTGCATCAAGCTTGGCTGGAGCCATCCCTGTGGAATTTCAGCA	1753
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	900
Db	1754	ACACTGTTAAGTTGACGTCGGGTCTGAAAGTGTAGAGTGAAGATGGAAAATTGCAGT	1813
Qy	901	TGAAGGGAACAACATATGGGTATGCTCAAAAGCATTCAAATTGCTAGGACTCCGCTG	960
Db	1814	TGAAGGGAACAACCTACGGCGTCTGTTCAAAGGCTTCAAGTTCTGGACTCCGCAG	1873
Qy	961	ACACTGGTCATGGAACGGTGGTCTGGAACTGCAGTATAACGGAAAAGACGGCCTTGCA	1020
Db	1874	ACACAGGTACGGCACTGTGGTGTGGAATTGCAGTACACTGGCACGGATGGACCTTGCA	1933
Qy	1021	AAGTGCCCATTCTCTGTGGCTTCCCTGAACGACCTAACCCGTTGGAAGGCTGGTA	1080
Db	1934	AAGTTCCCATTCTCGTCAGTGGCTTCCCTGAACGACCTAACACCGGTGGCAGATTGTCA	1993
Qy	1081	CTGTGAATCCATTGTGTCTGGCTACGGCAAACCGAAGGTTTGATTGAACCTGAAC	1140
Db	1994	CTGTTAACCTTTGTTCAAGTGGCCACGGCCAATGCCAAGGTCTGATTGAACCTGGAA	2053
Qy	1141	CCCCGTTAGTGACTCTTACATCGTGGTGGGGAGAGGGAGAACAGCAGATAAACCAACT	1200
Db	2054	CACCTTTGGAGACTCATACATAGTGGTAGGCAGAGGGAGAACACAGATCAATCACCATT	2113
Qy	1201	GGCACAAATCTGGGAGCAGTATTGAAAGGCTTCACCACTACACTCAGAGGAGCTAAC	1260
Db	2114	GGCATAAGTCTGGAAGCAGCATCGGCAAAGCCTTACAACCACTCTCAAAGGGCGCAGA	2173
Qy	1261	GACTTGAGCTTGGAGACACTGCCTGGATTGGATCAGTCGGAGGGTTTCACCT	1320
Db	2174	GATTAGCCGCTCTAGGAGACACAGCTGGGACTTGGATCAGTTGGAGGGTGTTCACCT	2233

Qy	1321 CGGTAGGGAAAGCCATACACCAAGTTTGGAGGAGCCTTAGATCACTCTTGGAGGGA 1380 
Db	2234 CAGTAGGGAAAGGCTGTCCATCAAGTGTTCGGTGGAGCGTTCCGCTCACTGTTGGAGGTA 2293
Qy	1381 TGTCCCTGGATCACACAGGGGCTTCTGGGAGCTCTGCTGTGGATGGAAATTAAACGCC 1440 
Db	2294 TGTCCCTGGATAACGCAGGGATTGCTGGGGCTTCTGTTGTGGATGGCATCAATGCTC 2353
Qy	1441 GTGACAGGTCAATTGCTATGACGTTCCCTGGGGTTGGAGGAGTCTTGCTCTCCTTCGG 1500 
Db	2354 GTGACAGGTCCATAGCTCTCACGTTCTCGCAGTTGGAGGAGTCTGCTCTCCTCTCCG 2413
Qy	1501 TCAACGTCCATGCTG 1515 
Db	2414 TGAACGTGCACGCTG 2428

## RESULT 10

AY262283

LOCUS AY262283 10984 bp RNA linear VRL 29-OCT-2003  
 DEFINITION West Nile virus isolate KN3829 polyprotein gene, complete cds.  
 ACCESSION AY262283  
 VERSION AY262283.1 GI:30230630  
 KEYWORDS .  
 SOURCE West Nile virus (WNV)  
 ORGANISM West Nile virus  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 Flavivirus; Japanese encephalitis virus group.  
 REFERENCE 1 (bases 1 to 10984)  
 AUTHORS Charrel,R.N., Brault,A.C., Gallian,P., Lemasson,J.-J., Murgue,B.,  
 Murri,S., Pastorino,B., Zeller,H., de Chesse,R., de Micco,P. and de  
 Lamballerie,X.  
 TITLE Evolutionary relationship between Old World West Nile virus  
 strains. Evidence for viral gene flow between africa, the middle  
 east, and europe  
 JOURNAL Virology 315 (2), 381-388 (2003)  
 PUBMED 14585341  
 REFERENCE 2 (bases 1 to 10984)  
 AUTHORS Brault,A.C. and de Lamballerie,X.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-MAR-2003) Division of Vector-Borne Infectious  
 Diseases, Centers for Disease Control and Prevention, P.O. Box  
 2087, Fort Collins, CO 80522, USA  
 FEATURES Location/Qualifiers  
 source 1. .10984  
 /organism="West Nile virus"  
 /mol\_type="genomic RNA"  
 /isolate="KN3829"  
 /specific\_host="Culex univittatus"  
 /db\_xref="taxon:11082"  
 5'UTR 1. .60  
 CDS 61. .10362  
 /codon\_start=1  
 /product="polyprotein"  
 /protein\_id="AAP20887.1"  
 /db\_xref="GI:30230631"  
 /translation="MSKKPGGPBKSRVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI"

RFVLALLAFAFRFTAIAPIRAVLDWRGVNKQTAMKHLLSFKKELGTITSAINRRSSKQ  
KKRGGGNTGIAAMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR  
AMDGVMCDDTITYECVPLSAGNDPEDIDCWCTKSAVVYVRYGRCTKTRHSRSRRSLT  
VQTHGESTLANKKGAWMDSTKATTRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV  
VFVVLVLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK  
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRGVVDRGWG  
NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIG  
ATQAGRFSITPAAPSYTLLKIGEYGEVTVDCEPRSGIDTNAYYMTVGTKTFLVHREW  
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSIALGSQEGALHQALAGAIPEFSS  
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCASKFKFLGTPADTGHTVVLELOYTGTDG  
PCKVPISSVASLNDLTPVGRLVTVNPVFSVATANAKVILIELEPPFGDSYIVVGRGEQQ  
INHHWHKSGSSIGKAFTTLLKGQAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF  
RSLFGGMWSITQGLLALLWMGINARDSTIALTFLAVGGVLLFLSVNVHADTGCAID  
ISRQELRCGSGVFIHNNDVEAWMDRYKYYPETPQGLAKIIQKAHEGVCGRLRSVSRL  
QMWEVKDELNTLLKENGVDLSVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSL  
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECD  
KIIGTAVKNNLAIHSDSLWYIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES  
DLIIPVTLAGPRSNHNRPGYKTQNQGPWDEGRVEIDFDYCPGTVTLSESCGRHPA  
TRTTTESGKLITDWCCRSCTLPPRLYQTDSCGCWYGMERPQRHDEKTLVQSQVNAYNA  
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYADVLRYVIL  
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTQNENILLMLAAVFFQMA  
DARQILLWEIPDVLSLAVAWMILRAITFTTSNVVVPLALLTPGLRCLNLDVYRIL  
LLMVGIGSIREKRSAAKKKGASLLCLALASTGLFNPMLAAGLIACDPNRKRGWPA  
TEVMTAVGLMFAIVGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDWMIERTADIS  
ESDAEITGSERVDVRLDDGNFQLMDPGAPWKIWMLRMACLAISAYTPWAILPSVV  
GFWITLQYTKRGGVWLWDTPSPKKEYKKGDTTGVYRIMTRGLLGSYQAGAGVMEGVFH  
TLWHTTKGAALMSGEGRLLDPYWSVKEDRLCYGGPWKLQHKWNGQDEVQMIIVEPGK  
VKNVQTKPGVKTPGEIGAVTLDFTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS  
AIVQGERMDEPIPAGFEPEMPLRKQITVLDLHPGAGKTRRILPQIKEAINRLRTAV  
LAPTRVVAEAEALRGLPIRYQTSAVTREHNGNEIVDVMCHATLTHRLMSPHRVPNY  
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAFMTATPPGTSDFPESNSPISDLQ  
TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKVVQLNRKSYETEY  
PKCKNDDDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTA  
SAAQRRGRIGRNPSPQVGDECYCGHTNEDDSNFAHWTEARIMLDNNINMPNGLIAQFYQ  
PEREKVYTMGEYRLGEERKNFELLRTADLPVWLAYKVAAGVSYHDRRCFDGPR  
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDASGKRSQIGLIEV  
LGKMPFHFMGKTWEALDTMYVWATAEKGGRAHRMAEEELPDALQTIALLSVMTMG  
VFFLLMQRKGIGKIGLGGVVLGVATFFCWMAEVPGTKIAGMLLSLLMIVLIPPEK  
QRSQTDNQLAFLICVMTLVSABAANEMGWLDTKSDISSLFGQRIEVKENFSMGEFL  
LDLRPATAWSLYAVTTAVLTPLLKLITSVDINTSLTSINVQASALFTLARGFPFVDV  
GVSALLAAGCWGQVTLTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA  
VVDGIVATDPELERTPIMQKKVGQIMLILVSLAAVVNPVSKTVREAGILITAAV  
TLWENGASSVWNATTAGLCHIMRGWLSCLSITWTLIKNDMDKPLKRGAKGRTLGE  
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNGVTGGHPVSRGTAKLRLV  
RFLEPGVKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGNIVTM  
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL  
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSASGNVVHSVNMTSQVLLG  
RMEKRTWKGPQYEEVDNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH  
PYRTWNYHGSYDVKPTGSASSLNGVVRLLSKPWDITNVTTMAMTDTPFGQQRVFK  
EKVDTKAEPPEGVKYVLNETTNLWAFLAREKRPRMCREEFIRKVNNAALGAMFE  
EQNQWRSAREAVEADPKFWEMVDEEREALRGECHTCIYNMMGKREKKPGEFGKAKGSR  
AIWFMWL GARFLEFEALGFLNEDHWLGRNSGGGVEGLGLQKLG YI LREVGTRPGGKI  
YADDTAGWDTRITRADLEAKVLELLDGEHRRRLARAIIELTYRHKVVKVMRPAADGR  
TVMDVI SREDQRGSGQVVYTAINTFTNLA VQLVRMMEGEGVIGPDDVEKLTKGKPKV  
RTWL FENCEERLSRMAVGDDCVVKPLDDRATSLHFLNAMS KVRKDIQEWPSTG  
DWQQVPCNSHFTELIMKDGR TLVPCRGQDELVGRARISPAGWNVRDTACLA  
KSYA QMWLLLYFHRRDLRLMANAICSAVPNVWPTGRTTW SI HAGGEWT  
TEDMLEVWN RVW

IEENEWMEDKTPVEKWSDPYSGKREDIWCGLIGTRARATWAENIQVAINQVRAIIG  
DEKYVDYMSSLKRYEDTTLVEDTVA"  
mat\_peptide 61. .429 /product="capsid"  
mat\_peptide 430. .705 /product="prM"  
mat\_peptide 706. .930 /product="M"  
mat\_peptide 931. .2433 /product="envelope"  
mat\_peptide 2434. .3492 /product="NS1"  
mat\_peptide 3493. .4182 /product="NS2A"  
mat\_peptide 4183. .4575 /product="NS2B"  
mat\_peptide 4576. .6432 /product="NS3"  
mat\_peptide 6433. .6879 /product="NS4A"  
mat\_peptide 6880. .7644 /product="NS4B"  
mat\_peptide 7645. .10359 /product="NS5"  
3'UTR 10363 10984

ORIGIN

Query Match 65.8%; Score 999.8; DB 10; Length 10984;  
Best Local Similarity 78.7%; Pred. No. 0;  
Matches 1193; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

Qy	1	CGGAATTCACTGTTAGGAATGAGCAACAGGGACTTCCTGGAGGGAGTGTCTG	60
Db	920	CAGCCTACAGCTCAACTGTCTTGAATGAGCAACAGAGACTTCTGGAGGGAGTATCTG	979
Qy	61	GAGCTACATGGGTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAAG	120
Db	980	GAGCAACATGGGTGGATTGGTCTCGAAGGCGACAGCTGCGTGACTATCATGTCTAAGG	1039
Qy	121	ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAATCTCGCAGATGTGC	180
Db	1040	ACAAGCCTACCATTGATGTGAAGATGATGAATATGGAGGCTGCCAACCTGGCAGAGGTCC	1099
Qy	181	GTA GCT ACT GCT ACT TAGCTCGGT CAGT GAT CT GT CAAC AAA AGCC CGT GTCCA ACCA	240
Db	1100	GCAGTTATTGCTATTGGCTACCGTCAGCGATCTCCACCAAAGCTGCATGCCCGACCA	1159
Qy	241	TGGGTGAAGCTCACAAACGAGAAAAGAGCCGACCCCTGCCTTGTGCAAGCAAGGCGTCG	300
Db	1160	TGGGAGAAGCTCACAAATGACAAACGTGCTGACCCAGCTTTGTGCAAGACAAGGAGTGG	1219
Qy	301	TAGACAGAGGATGGGGAATGGATGCGGACTGTTGGAAAGGGGAGCATTGACACATGTG	360
Db	1220	TGGACAGGGGCTGGGCAACGGCTGCGGACTATTGGCAAAGGAAGCATTGACACATGCG	1279
Qy	361	CAAAGTTGCCTGTACAACCAAGGCAACTGGTTGGATTATCCAGAAGGAAAACATCAAGT	420

Db	1280	CCAAATTGCCTGCTCCACCAAGGCAACAGGAAGAACCATCTTGAAGAGAATATCAAGT	1339
Qy	421	ACGAGGTTGCCATATTGTGCATGGCCCGACGACTGTGAAATCACATGGCAATTATTCAA	480
Db	1340	ATGAAGTGGCCATCTTGTCCATGGACCAACCACGTGGAGTCGCATGGAAACTACTCCA	1399
Qy	481	CACAGATAGGGCTACCCAAGCAGGAAGGTCAGCATAACTCCATCGGCACCATCCTACA	540
Db	1400	CACAGATTGGGCCACTCAGGCAGGGAGATTCAGCATCACTCCTGC GGCGCCTTCATACA	1459
Qy	541	CGCTGAAGTTGGGTGAGTATGGTGAGGTACAGTTGACTGTGAGGCCACGGTCAGGAATAG	600
Db	1460	CACTAAAGCTTGGAGAATATGGAGAAGTGACAGTGGACTGTGAACCACGGTCAGGGATTG	1519
Qy	601	ACACTAGCGCTTACTACGTTATGTCAGTGGGTGCGAAGTCCTTCTGGTTACCGAGAAT	660
Db	1520	ACACCAATGCTTACTACGTGATGACTGTTGAAACAAAGACGTTTGGTCCATCGTGAGT	1579
Qy	661	GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG	720
Db	1580	GGTCATGGACCTCAACCTCCCTGGAGCAGTGTGGAAAGTACTGTGTGGAGGAACAGAG	1639
Qy	721	AAACACTGATGGAGTTGAAGAACCTCATGCCACCAAAACAAATCTGTCGTAGCTCTAGGGT	780
Db	1640	AGACGTTAATGGAGTTGAGGAACCACACGCCACAAAGCAGTGTGTGATAGCATTGGGCT	1699
Qy	781	CGCAGGAAGGTGCCTTGCACCAAGCTCTGGCTGGAGCAATTCCCTGGTGGAGTTCTCAAGCA	840
Db	1700	CACAAGAGGGAGCTCGCATCAAGCTTGGCTGGAGCCATCCCTGTGGAAATTTCAGCAAGCA	1759
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	900
Db	1760	ACACTGTCAAGTTGACGTGGTGGTCAATTGAAGTGTAGAGTGAAGATGGAAAAATTGCAGT	1819
Qy	901	TGAAGGAAACAACATATGGTGTATGCTAAAAGCATTCAAATTGCTAGGACTCCGCTG	960
Db	1820	TGAAGGAAACAACCTACGGCGTCTGTTCAAAGGCTTCAAGTTCTGGACTCCGCAG	1879
Qy	961	ACACTGGTCATGGAACGGTGGTGGACTGCAGTATACCGGAAAAGACGGGCCTTGCA	1020
Db	1880	ACACAGGTACGGCACTGTGGTGGAAATTGCACTGGTACGGATGGACCTTGCA	1939
Qy	1021	AAGTGCCCATTTCTCTGTGGCTCCCTGAACGACCTACACCGTTGGAGGCTGGTGA	1080
Db	1940	AAGTCCCACATCTCGTCAAGTGGCTCATTGAACGACCTAACACCAGTGGCAGATTGGTCA	1999
Qy	1081	CTGTGAATCCATTGTGTCTGTGGCTACGGCCAACCTCGAAGGTTTGATTGAACCTCGAAC	1140
Db	2000	CTGTCAACCCTTTGTTCAAGTGGCCACGGCAATGCAAGGTCTGATTGAACCTGGAAC	2059
Qy	1141	CCCCGTTAGTGAACCTTACATCGTGGTGGAGAGGGAGAACAGCAGATAAACCAACT	1200
Db	2060	CACCCCTTGGAGACTCATACATAGTGGTAGGCAGAGGAGAACACAGATCAATCACCATT	2119
Qy	1201	GGCACAAATCTGGGAGCAGTATTGAAAGGCTTCACCAACTACACTCAGAGGAGCTAAC	1260
Db	2120	GGCATAAGTCTGGAAGCAGCATTGGCAAAGCCTTACAACCAACTCTCAAAGGGCGCAGA	2179

Qy	1261	GACTTGCAGCTCTGGAGACACTGCCTGGGATTTGGATCAGTCGGAGGGGTTTCACCT	1320
Db	2180	GATTAGCCGCTTAGGAGACACAGCTGGGACTTGATCAGTTGGAGGGGTGTTCACCT	2239
Qy	1321	CGGTAGGGAAAGCCATACACCAAGTTTGAGGAGCCTTAGATCACTTTGGAGGGA	1380
Db	2240	CAGTAGGGAAGGCTGCCATCAAGTGGTGGAGCATTCCGCTCACTGTTGGAGGTA	2299
Qy	1381	TGTCCTGGATCACACAGGGCTTCTGGAGCTCTGCTGTGGATGGAAATTAACGCC	1440
Db	2300	TGTCCTGGATAACGCAGGGATTGCTGGGGCTCTGTTGTGGATGGCATCAATGCTC	2359
Qy	1441	GTCACAGGTCAATTGCTATGACGTTCTGGCGGTGGAGGAGTCTGCTCTCCTTCGG	1500
Db	2360	GTCACAGGTCCATAGCTCTCACGTTCTCGCAGTTGGAGGAGTCTGCTCTCCTCTG	2419
Qy	1501	TCAACGTCCATGCTG	1515
Db	2420	TGAACGTGCACGCTG	2434

#### RESULT 11

AF404757

LOCUS AF404757 11029 bp ss-RNA linear VRL 23-JUL-2002  
 DEFINITION West Nile virus isolate WN Italy 1998-equine, complete genome.  
 ACCESSION AF404757  
 VERSION AF404757.1 GI:21929240  
 KEYWORDS .  
 SOURCE West Nile virus  
 ORGANISM West Nile virus  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 Flavivirus; Japanese encephalitis virus group.  
 REFERENCE 1 (bases 1 to 11029)  
 AUTHORS Lanciotti, R.S., Ebel, G.D., Deubel, V., Kerst, A.J., Murri, S.,  
 Meyer, R., Bowen, M., McKinney, N., Morrill, W.E., Crabtree, M.B.,  
 Kramer, L.D. and Roehrig, J.T.  
 TITLE Complete genome sequences and phylogenetic analysis of West Nile  
 virus strains isolated from the United States, Europe, and the  
 Middle East  
 JOURNAL Virology 298 (1), 96-105 (2002)  
 PUBMED 12093177  
 REFERENCE 2 (bases 1 to 11029)  
 AUTHORS Deubel, V., Bowen, M., Meyer, R., McKinney, N. and Morrill, W.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-AUG-2001) Division of Vector-Borne Infectious  
 Diseases, Centers for Disease Control & Prevention, Rampart Road,  
 Fort Collins, CO 80521, USA  
 FEATURES Location/Qualifiers  
 source 1. .11029  
 /organism="West Nile virus"  
 /mol\_type="genomic RNA"  
 /isolate="WN Italy 1998-equine"  
 /specific\_host="equine"  
 /db\_xref="taxon:11082"  
 /country="Italy"  
 CDS 97. .10398

/note="contains capsid, pre-membrane, envelope, NS1, NS2a, NS2b, NS3, NS4a, NS4b, and NS5"  
/codon\_start=1  
/product="polyprotein precursor"  
/protein\_id="AAM81753.1"  
/db\_xref="GI:21929241"  
/translation="MSKKPGGPGKSRAVNMLKRGMPRVSLIGLKRAMLSLIDGKGPI  
RFVLALLAFRFTAIAPTRAVLDRWGVNKQTAMKHLLSFKKEGLTTSAINRRSSKQ  
KKRGGKTGIAVMIGLIASVGAVTLSNFQGVMMTVNATDVTDVITIPTAAGKNLCIVR  
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAYVRYGRCTKTRHSRRSRRSLT  
VQTHGESTLANKGAWMDSTKATTRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV  
VFVVLVLLVAPAYSFNCLGMSNRDFLEGVSATWVDLVLEGDSCVTIMSKDKPTIDVK  
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG  
NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIG  
ATQAGRFSITPAAPSYTTLKGEYGEVTVDCEPRSGIDTNAYYMTVGTKTFLVHREW  
MDLNLPWSSAGSTVWRNRETLMEFEPHATKQSIALGSQEGALHQALAGAIPVEFSS  
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCASKFKFLGPADTGHGTVVLELQYTGTDG  
PCKVPISSVASLNDTPVGRILTVNPVSVATANAKVIELEPFFGDSYIVVGRGEQQ  
INHHWHKSGSIGKAFTTLKGQAQLRAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF  
RSLFGGMSWITQGLLGALLWMGINARDRIALTFLAVGGVLLFLSVNVHADTGCAID  
ISRQEELRCGSGVFIHNDEAVMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVRLEH  
QMWEVKDELNTLLKENGVDLSVVEKQEGMYKSAPKRLTATTTEKLEIGWKAWGKSIL  
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECD  
KIIGTAVKNLAIHSDSLWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDI  
LES DLIIPVTLAGPRSNHNRPGYKTQNGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA  
TRTTTESGKLITDWCCRSCTLPPPLYQTDSCGCWYGMIEPQRHDEKTLVQSQVNAYNA  
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAI LIA LLVLFVGITYTDVLRVIL  
GAAFAESNSGGDVVHLMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMA  
YH DARQILLWEIPDVLSLAVAWMIRAITFTTSNVVPLLLALTPGLRCLNDVYRIL  
LLMVGIGSLIREKRSAAKKKGASLLCLALASTGLFNPMLAAGLIACDPNRKRGWPA  
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDWIERTADIS  
ESDAEITGSSSERVDVRLDDDGPNQLMDPGAPWKIWMLRMACLAISAYTPWAI  
LPSVI GFWITLQYTKRGGVLDTPSPKEYKKGDTTGVRIMTRGLLGSYQAGAGVMVEGVFH  
TLWHTTKGAALMSGEGRRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMI  
VVEPGKNVKNVQTCKPGVFKTPEGEIGAVTLDPTGTSGSPIVDKNGDVIGLY  
GNGVIMPNGSYIAS AIVQGERMDEPI PAGFEPEMLRKQITVLDLHPGAGKTRRIL  
PQIIKEAINRRLRTAV LAPTRVVAEAEALRGLPIRYQTSAVTREHNGNE  
IIVDVMCHATLTHRLMSPHRPVNY NLFVMDEAHFTDPASIAARGYISTK  
VELGEAAAIFMTATPPGTSDPFPESNSPISDLQ TEIPDRAWNSGYEWITEY  
IGKTVWFVPSVKMGNEIALCLQRAGKVVQLNRKS  
YETEY PKCKNDDWDFTT  
DI SEMGANFKASRVIDSRKS  
VKPTI ITEGEGRVIL  
GEPSAVTAA SAAQRRGRIGRNP  
S QVGDEYCYGGHTNE  
EDDSNFAHWT  
EARIMLDNIN  
NMPNGLIAQFYQ PEREKVY  
TMDGEYRL  
GEERKNF  
LELLRT  
ADLPV  
WLA  
YK  
VAAAG  
SYH  
D  
RRWC  
FDGPR TNTILED  
NN  
NE  
EV  
E  
V  
IT  
K  
L  
G  
E  
R  
K  
I  
L  
R  
P  
W  
I  
D  
A  
R  
V  
Y  
S  
D  
H  
Q  
A  
L  
K  
A  
F  
K  
D  
F  
A  
S  
G  
K  
R  
S  
Q  
I  
G  
L  
I  
E  
V LGKMPEHFMGKT  
WEALDT  
MYV  
VATA  
AEKG  
G  
RA  
H  
R  
M  
A  
E  
L  
P  
D  
A  
L  
Q  
T  
I  
A  
L  
L  
S  
V  
M  
T  
M  
G VF  
F  
L  
L  
M  
Q  
R  
K  
G  
I  
G  
K  
I  
L  
G  
G  
V  
V  
L  
G  
V  
A  
T  
F  
C  
W  
M  
A  
E  
V  
P  
G  
T  
K  
I  
A  
G  
M  
L  
L  
S  
L  
L  
M  
I  
V  
L  
I  
P  
E  
P  
E  
K  
Q  
R  
S  
Q  
T  
D  
N  
Q  
L  
A  
V  
F  
L  
I  
C  
V  
M  
T  
L  
V  
S  
A  
V  
A  
N  
E  
M  
G  
W  
L  
D  
K  
T  
K  
S  
D  
I  
S  
S  
L  
F  
G  
Q  
R  
I  
E  
V  
K  
E  
N  
F  
S  
M  
G  
E  
F  
L  
L  
D  
R  
P  
A  
T  
A  
W  
S  
L  
Y  
A  
V  
T  
T  
A  
V  
L  
T  
P  
L  
L  
K  
H  
L  
I  
T  
S  
D  
Y  
I  
N  
T  
S  
L  
T  
S  
I  
N  
V  
Q  
A  
S  
A  
F  
T  
L  
A  
R  
G  
F  
P  
F  
D  
V  
G  
V  
S  
A  
L  
L  
A  
A  
G  
C  
W  
Q  
V  
T  
L  
T  
V  
T  
A  
T  
L  
L  
F  
C  
H  
Y  
A  
M  
V  
P  
G  
W  
Q  
A  
E  
A  
M  
R  
S  
A  
Q  
R  
R  
T  
A  
G  
I  
M  
K  
N  
A  
V  
V  
D  
G  
I  
V  
A  
T  
D  
V  
P  
E  
L  
R  
T  
P  
I  
M  
Q  
K  
V  
G  
Q  
I  
M  
L  
I  
L  
V  
S  
L  
A  
A  
V  
V  
V  
N  
P  
S  
V  
K  
T  
V  
R  
E  
A  
G  
I  
L  
I  
T  
A  
A  
V  
T  
L  
W  
E  
N  
G  
A  
S  
S  
V  
W  
N  
A  
T  
T  
A  
I  
G  
L  
C  
H  
I  
M  
R  
G  
G  
W  
L  
S  
C  
L  
S  
I  
T  
W  
T  
L  
I  
K  
N  
M  
D  
K  
P  
G  
L  
K  
R  
G  
G  
A  
K  
R  
T  
L  
G  
E  
V  
W  
K  
E  
R  
L  
N  
Q  
M  
T  
K  
E  
E  
F  
T  
R  
Y  
R  
K  
E  
A  
I  
I  
E  
V  
D  
R  
S  
A  
A  
K  
H  
A  
R  
K  
E  
G  
N  
V  
T  
G  
G  
H  
P  
V  
S  
R  
G  
T  
A  
K  
L  
R  
W  
L  
V  
R  
E  
F  
L  
P  
E  
V  
G  
K  
V  
I  
D  
L  
G  
C  
G  
R  
G  
G  
W  
C  
Y  
Y  
M  
A  
T  
Q  
K  
R  
V  
Q  
E  
V  
R  
G  
Y  
T  
K  
G  
G  
P  
G  
H  
E  
E  
P  
Q  
L  
V  
Q  
S  
Y  
G  
W  
N  
I  
V  
T  
M  
G  
D  
V  
F  
Y  
R  
P  
S  
E  
C  
C  
D  
T  
L  
L  
C  
D  
I  
G  
E  
S  
S  
S  
A  
E  
V  
E  
E  
H  
R  
T  
I  
R  
V  
L  
E  
M  
V  
E  
D  
W  
L  
H  
R  
G  
P  
R  
E  
F  
C  
V  
K  
V  
C  
P  
Y  
M  
P  
K  
V  
I  
E  
K  
M  
E  
L  
L  
Q  
R  
R  
Y  
G  
G  
L  
V  
R  
N  
P  
L  
S  
R  
N  
S  
T  
H  
E  
M  
Y  
W  
V  
S  
R  
A  
G  
N  
V  
V  
H  
S  
V  
N  
M  
T  
S  
Q  
V  
L  
L  
R  
E  
Y  
S  
S  
T  
W  
H  
D  
E  
N  
H  
P  
Y  
R  
T  
W  
N  
Y  
H  
G  
S  
Y  
D  
V  
K  
P  
T  
G  
S  
A  
S  
L  
V  
N  
G  
V  
V  
R  
L  
L  
S  
K  
P  
W  
D  
T  
I  
T  
N  
V  
T  
T  
M  
A  
M  
T  
D  
T  
P  
F  
G  
Q  
Q  
R  
V  
F  
K  
E  
V  
D  
T  
K  
A  
P  
E  
P  
P  
E  
G  
V  
K  
Y  
V  
L  
N  
E  
T  
T  
N  
W  
L  
W  
A  
F  
L  
A  
R  
E  
K  
R  
P  
R  
M  
C  
S  
R  
E  
E  
F  
I  
R  
K  
V  
N  
S  
A  
A  
L  
G  
A  
M  
F  
E

EQNQWRSAREAVEDPKFWEMVDEEREALRGECHTCIYNMMGKREKKPGEGKAKGSR  
AIWFMLWLGRARFLEFEALGFLNEDHWLGRNSGGGVVEGLGLQKLGYILREVGTRPGGKI  
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLLARAIIELTYRHKKVVMRPAADGR  
TVMDVISREDQRGSQVVTYALNTFTNLAVQLVRMMEGEVIGPDDVEKLTKGKGPKV  
RTWLFENGEERLSRMAVGDDCVVKPLDDRATSLHFLNAMSVKRKDIQEWPSTGWY  
DWQQVPFCSNHFTELIMKDRTLUVPCRGQDELVGRARISPAGWNVRDTACLAKSYA  
QMWLLLYFHRRDLRLMANAICSAVPVNWPTGRRTTWSIHAGGEWMTTEDMLEVNRVW  
IEENEWMEDKTPVKEWSDVPSGKREDIWCGLIGTRARATWAENIQVAINQVRAIIG  
DEKYVDYMSLKRVEDTTLVEDTBL"

## ORIGIN

Db	1556	ACACTAATGCCCTACTACGTGATGACTGTTGGAACAAAGACGTTGGTCCATCGTGAGT	1615
Qy	661	GGTTTATGGACCTGAAACCTTCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG	720
Db	1616	GGTCATGGACCTCAACCTCCCTGGAGCAGTGTGGAAGTACTGTGTGGAGGAACAGAG	1675
Qy	721	AAACACTGATGGAGTTGAAGAACCTCATGCCACCAAACAATCTGCGTAGCTAGGGT	780
Db	1676	AGACGTTGATGGAGTTGAGGAACCACGCCACAAAGCAGTGTGATAGCATTGGGCT	1735
Qy	781	CGCAGGAAGGTGCCTTGCACCAAGCTCTGGCTGGAGCAATTCCCTGGAGTTCTCAAGCA	840
Db	1736	CACAAGAGGGAGCTGCAAGCTTGGCTGGAGCCATCCCTGTGGAATTTCAGCAAGCA	1795
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	900
Db	1796	ACACTGTCAAGTTGACGTCGGGCATTGAGTGTAGAGTGAAGATGGAAAAATTGCAGT	1855
Qy	901	TGAAGGAAACAACATATGGTGTATGCTAAAAGCATTCAAATTGCTAGGACTCCGCTG	960
Db	1856	TGAAGGAAACAACATTACGGCGTCTGTTCAAAGGCTTCAAGTTCTGGGACTCCGCAAG	1915
Qy	961	ACACTGGTCATGGAACGGTGGTGGACTGCAGTATAACCGGAAAAGACGGGCCTTGCA	1020
Db	1916	ACACAGGTACGGCACTGTGGTGGAAATTGAGTACACTGGCACGGATGGACCTTGCA	1975
Qy	1021	AAGTGCCCATTCTCTGTGGCTCCCTGAACGACCTACACCGGTTGGAAGGCTGGTGA	1080
Db	1976	AAGTTCCCATCTCGTCAAGTGGCTTCATTGAACGACCTAACACCGGTAGGCAGATTGGTCA	2035
Qy	1081	CTGTGAATCCATTGTTGCTGTGGCTACGGCCAACCTCGAAGGTTTGATTGAACCGAAC	1140
Db	2036	CTGTTAACCTTTGTTCAAGTGGCACGGCAACGCAAGGTCTGATTGAACCTGGAAC	2095
Qy	1141	CCCCGTTAGTGACTCTTACATCGTGGTGGGGAGAGGAGAACAGCAGATAAACCAACT	1200
Db	2096	CACCTTTGGAGACTCATACATAGTGGTAGGCAGAGGAGAACACAGATCAATCACCATT	2155
Qy	1201	GGCACAAATCTGGGAGCAGTATTGAAAGGCTTCACCACTACACTCAGAGGAGCTAAC	1260
Db	2156	GGCATAAGTCTGGAAGCAGCATTGGCAAAGCCTTACAACCACTCTCAAAGGGCGCAGA	2215
Qy	1261	GACTTGCAGCTTGGAGACACTGCCTGGATTTGGATCAGTCGGAGGGGTTTCACCT	1320
Db	2216	GATTAGCCGCTCTAGGAGACACAGCTTGGACTTCGGATCAGTTGGAGGGGTGTTACCT	2275
Qy	1321	CGGTAGGGAAAGCCATACACCAAGTTTGGAGGAGCCTTAGATCACTCTTGGAGGGA	1380
Db	2276	CAGTAGGGAAAGGCTGCCATCAAGTGTCCGGAGCATTCCGCTCACTGTTGGAGGTA	2335
Qy	1381	TGTCCTGGATCACACAGGGCTTCTGGAGCTCTGCTGTGGATGGAAATTACGCC	1440
Db	2336	TGTCCTGGATAACGCAGGGATTGCTGGGGCTCTCTGTTGGATGGCATCAATGCTC	2395
Qy	1441	GTGACAGGTCAATTGCTATGACGTTCTGGCGGGTGGAGGAGTCTGCTCTCCTTCGG	1500
Db	2396	GTGACAGGTCCATAGCTCACGTTCTCGCAGTTGGAGGAGTCTGCTCTCCTCTCCG	2455

Qy 1501 TCAACGTCCATGCTG 1515  
| | | | | | | |  
Db 2456 TGAACGTGCACGCTG 2470

RESULT 12

AY660002

LOCUS AY660002 11029 bp RNA linear VRL 19-DEC-2004  
DEFINITION West Nile virus isolate Mex03 from Mexico, complete genome.  
ACCESSION AY660002  
VERSION AY660002.1 GI:55975602  
KEYWORDS .  
SOURCE West Nile virus (WNV)  
ORGANISM West Nile virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
Flavivirus; Japanese encephalitis virus group..  
REFERENCE 1 (bases 1 to 11029)  
AUTHORS Beasley,D.W.C., Davis,C.T., Estrada-Franco,J., Navarro-Lopez,R.,  
Campomanes-Cortes,A., Tesh,R.B., Weaver,S.C. and Barrett,A.D.T.  
TITLE Genome Sequence and Attenuating Mutations in West Nile Virus  
Isolate from Mexico  
JOURNAL Emerging Infect. Dis. 10 (12), 2221-2224 (2004)  
REFERENCE 2 (bases 1 to 11029)  
AUTHORS Beasley,D.W.C., Davis,T., Estrada-Franco,J.G., Tesh,R.B.,  
Weaver,S.C. and Barrett,A.D.T.  
TITLE Direct Submission  
JOURNAL Submitted (18-JUN-2004) Pathology, University of Texas Medical  
Branch, 301 University Blvd., Galveston, TX 77555-0609, USA  
FEATURES Location/Qualifiers  
source 1. .11029  
/organism="West Nile virus"  
/mol\_type="genomic RNA"  
/strain="TM171-03"  
/isolate="Mex03"  
/db\_xref="taxon:11082"  
/country="Mexico"  
CDS 97. .10398  
/codon\_start=1  
/product="polyprotein precursor"  
/protein\_id="AAV68177.1"  
/db\_xref="GI:55975603"  
/translation="MSKKPGGPDKSRAVNMLKRGMPRVLSLIGLKRAMLSLDGKGPI  
RFVLALLAFLRFTAIAPTRAVLDRWRGVNKQTAMKHLSFKKEGLTTSAINRRSSKQ  
KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR  
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAYVRYGRCKTRHSRRSRRSLT  
VQTHGESTLANKKGAWMDSTKATTRYLVKTESWILRNPGYALVAATGWMLGSNTMQRV  
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDCVTIMSKDKPTIDVK  
MMMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG  
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYPTQVG  
ATQAGRFSITPAAPSYTTLKLGEGYGEVTVDCEPRSGIDTNAYYMTVGTKTFLVHREW  
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAI PVEFSS  
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCNSKAFKFLGTPADTGHTVVLELQYTGTDG  
PCKVPISSVASLNDLTPVGRLLTVNPVFSVATANAKVILIELEPFFGDSYIVVGRGEQQ  
INHHWHKSGSSSIGKAFTTTLKGQAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF  
RSLFGGMSWITQGLLGALLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID  
ISRQEELRCGSGVFIHN DVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRL  
EH

QMWEAVKDELNTLLKENGVDLSVVEKQEGMYKSAPKRLTATTEKLEIGWKAWSKISL  
FAPELANNTFVVDGPETKECPTQNRAWSLEVEDFGFGLTSTRMFLKVRESNTTECD  
KIIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCWTPEHTLWGDI  
LESDLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGH  
RGPA  
TRTTTESGKLITDWCCRSCTLPPRLRYQTDGCGWYGMERIPQRHDEKTLVQS  
QVNAYNA  
DMIDPFQLGLLVVFATQEVLRKRWTAKISM  
PAILIAALLVLVFGITYTDVLRVIL  
GAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAFFQ  
MAYH  
DARQILLWEIPDVLSLAVAWMILRAITFTTSNVVVPLALLTPGLRCLNLDV  
YRIL  
LLMVGIGSLIREKRSAAAKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRG  
WP  
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTD  
MWIERTADIS  
W  
ESDAEITGSERVDVRLLDDGNFQLMDPGAPWKIWMRMVCLA  
ISAYTPWAI  
LP  
SVV  
GFWITLQYT  
KRG  
GV  
LWDT  
P  
S  
KEY  
KK  
G  
DT  
T  
G  
V  
Y  
R  
IM  
T  
R  
G  
L  
S  
Y  
Q  
A  
G  
A  
G  
V  
M  
E  
G  
V  
F  
H  
TLWHTTKGA  
ALMS  
GE  
G  
R  
L  
D  
P  
Y  
W  
G  
S  
V  
K  
E  
D  
R  
L  
C  
Y  
G  
G  
P  
W  
K  
L  
Q  
H  
K  
W  
N  
G  
Q  
D  
E  
V  
Q  
M  
I  
V  
V  
E  
P  
G  
K  
N  
V  
K  
N  
Q  
T  
K  
P  
G  
V  
F  
K  
T  
P  
E  
G  
E  
I  
G  
A  
V  
T  
L  
D  
P  
T  
G  
S  
P  
I  
V  
D  
K  
N  
G  
D  
V  
I  
G  
L  
Y  
G  
N  
G  
V  
I  
M  
P  
N  
G  
L  
I  
A  
Q  
F  
Y  
Q  
P  
E  
R  
E  
K  
V  
Y  
T  
M  
D  
G  
E  
Y  
R  
L  
R  
G  
E  
E  
R  
K  
N  
F  
L  
E  
L  
L  
R  
T  
A  
D  
L  
P  
V  
W  
L  
A  
Y  
K  
V  
A  
A  
G  
V  
S  
Y  
H  
D  
R  
R  
C  
F  
D  
G  
P  
R  
T  
N  
T  
I  
L  
E  
D  
N  
N  
E  
E  
V  
I  
T  
K  
L  
G  
E  
R  
K  
I  
L  
R  
P  
R  
W  
I  
D  
A  
R  
V  
Y  
S  
D  
H  
Q  
A  
L  
K  
A  
F  
K  
D  
F  
A  
S  
G  
K  
R  
S  
Q  
I  
G  
L  
I  
E  
V  
L  
G  
K  
M  
P  
E  
H  
F  
M  
G  
K  
T  
W  
E  
A  
L  
D  
T  
M  
Y  
V  
V  
A  
T  
A  
E  
K  
G  
G  
R  
A  
H  
R  
M  
A  
L  
E  
E  
L  
P  
D  
A  
L  
Q  
T  
I  
A  
L  
I  
A  
L  
S  
V  
M  
T  
G  
V  
F  
L  
L  
M  
Q  
R  
K  
G  
I  
G  
K  
I  
G  
L  
G  
G  
A  
V  
L  
G  
V  
A  
T  
F  
C  
W  
M  
A  
E  
V  
P  
G  
T  
K  
I  
A  
G  
M  
L  
L  
S  
L  
L  
M  
I  
V  
L  
I  
P  
E  
P  
K  
E  
Q  
R  
S  
Q  
T  
D  
N  
Q  
L  
A  
V  
F  
L  
I  
C  
V  
M  
T  
I  
V  
S  
A  
A  
N  
E  
M  
G  
W  
L  
D  
K  
T  
K  
S  
D  
I  
S  
L  
F  
G  
Q  
R  
I  
E  
V  
K  
E  
N  
F  
S  
M  
G  
E  
F  
L  
L  
D  
L  
R  
P  
A  
T  
A  
W  
S  
L  
Y  
A  
V  
T  
A  
V  
L  
T  
P  
L  
L  
K  
H  
L  
I  
T  
S  
D  
Y  
I  
N  
T  
S  
L  
T  
S  
I  
N  
V  
Q  
A  
S  
A  
L  
F  
T  
L  
A  
R  
G  
F  
P  
F  
V  
D  
V  
G  
S  
A  
L  
L  
A  
A  
G  
C  
W  
G  
Q  
V  
T  
L  
T  
V  
T  
A  
A  
T  
L  
L  
F  
C  
H  
Y  
A  
M  
V  
P  
G  
W  
Q  
A  
E  
A  
M  
R  
S  
A  
Q  
R  
R  
T  
A  
G  
I  
M  
K  
N  
A  
V  
V  
D  
G  
I  
V  
A  
T  
D  
V  
P  
E  
L  
R  
T  
P  
I  
M  
Q  
K  
V  
G  
Q  
I  
M  
L  
I  
L  
V  
S  
I  
A  
V  
V  
V  
N  
P  
S  
V  
K  
T  
V  
R  
E  
A  
G  
I  
L  
I  
T  
A  
A  
A  
V  
T  
L  
W  
E  
N  
G  
A  
S  
S  
V  
W  
N  
A  
T  
A  
I  
G  
L  
C  
H  
I  
M  
R  
G  
G  
W  
L  
S  
C  
L  
S  
I  
T  
W  
T  
L  
V  
K  
N  
M  
E  
P  
G  
L  
K  
R  
G  
G  
A  
K  
R  
T  
L  
G  
E  
H  
R  
M  
A  
L  
E  
E  
L  
P  
D  
A  
L  
Q  
T  
I  
A  
L  
I  
A  
L  
S  
V  
M  
T  
G  
V  
F  
L  
L  
M  
Q  
R  
K  
G  
I  
G  
K  
I  
G  
L  
G  
G  
A  
V  
L  
G  
V  
A  
T  
F  
C  
W  
M  
A  
E  
V  
P  
G  
T  
K  
I  
A  
G  
M  
L  
L  
S  
L  
L  
M  
I  
V  
L  
I  
P  
E  
P  
K  
E  
Q  
R  
S  
Q  
T  
D  
N  
Q  
L  
A  
V  
F  
L  
I  
C  
V  
M  
T  
I  
V  
S  
A  
A  
N  
E  
M  
G  
W  
L  
D  
K  
T  
K  
S  
D  
I  
S  
L  
F  
G  
Q  
R  
I  
E  
V  
K  
E  
N  
F  
S  
M  
G  
E  
F  
L  
L  
D  
L  
R  
P  
A  
T  
A  
W  
S  
L  
Y  
A  
V  
T  
A  
V  
L  
T  
P  
L  
L  
K  
H  
L  
I  
T  
S  
D  
Y  
I  
N  
T  
S  
L  
T  
S  
I  
N  
V  
Q  
A  
S  
A  
L  
F  
T  
L  
A  
R  
G  
F  
P  
F  
V  
D  
V  
G  
S  
A  
L  
L  
A  
A  
G  
C  
W  
G  
Q  
V  
T  
L  
T  
V  
T  
A  
A  
T  
L  
L  
F  
C  
H  
Y  
A  
M  
V  
P  
G  
W  
Q  
A  
E  
A  
M  
R  
S  
A  
Q  
R  
R  
T  
A  
G  
I  
M  
K  
N  
A  
V  
V  
D  
G  
I  
V  
A  
T  
D  
V  
P  
E  
L  
R  
T  
P  
I  
M  
Q  
K  
V  
G  
Q  
I  
M  
L  
I  
L  
V  
S  
I  
A  
V  
V  
V  
N  
P  
S  
V  
K  
T  
V  
R  
E  
A  
G  
I  
L  
I  
T  
A  
A  
A  
V  
T  
L  
W  
E  
N  
G  
A  
S  
S  
V  
W  
N  
A  
T  
A  
I  
G  
L  
C  
H  
I  
M  
R  
G  
G  
W  
L  
S  
C  
L  
S  
I  
T  
W  
T  
L  
V  
K  
N  
M  
E  
P  
G  
L  
K  
R  
G  
G  
A  
K  
R  
T  
L  
G  
E  
H  
R  
M  
A  
L  
E  
E  
L  
P  
D  
A  
L  
Q  
T  
I  
A  
L  
I  
A  
L  
S  
V  
M  
T  
G  
V  
F  
L  
L  
M  
Q  
R  
K  
G  
I  
G  
K  
I  
G  
L  
G  
G  
A  
V  
L  
G  
V  
A  
T  
F  
C  
W  
M  
A  
E  
V  
P  
G  
T  
K  
I  
A  
G  
M  
L  
L  
S  
L  
L  
M  
I  
V  
L  
I  
P  
E  
P  
K  
E  
Q  
R  
S  
Q  
T  
D  
N  
Q  
L  
A  
V  
F  
L  
I  
C  
V  
M  
T  
I  
V  
S  
A  
A  
N  
E  
M  
G  
W  
L  
D  
K  
T  
K  
S  
D  
I  
S  
L  
F  
G  
Q  
R  
I  
E  
V  
K  
E  
N  
F  
S  
M  
G  
E  
F  
L  
L  
D  
L  
R  
P  
A  
T  
A  
W  
S  
L  
Y  
A  
V  
T  
A  
V  
L  
T  
P  
L  
L  
K  
H  
L  
I  
T  
S  
D  
Y  
I  
N  
T  
S  
L  
T  
S  
I  
N  
V  
Q  
A  
S  
A  
L  
F  
T  
L  
A  
R  
G  
F  
P  
F  
V  
D  
V  
G  
S  
A  
L  
L  
A  
A  
G  
C  
W  
G  
Q  
V  
T  
L  
T  
V  
T  
A  
A  
T  
L  
L  
F  
C  
H  
Y  
A  
M  
V  
P  
G  
W  
Q  
A  
E  
A  
M  
R  
S  
A  
Q  
R  
R  
T  
A  
G  
I  
M  
K  
N  
A  
V  
V  
D  
G  
I  
V  
A  
T  
D  
V  
P  
E  
L  
R  
T  
P  
I  
M  
Q  
K  
V  
G  
Q  
I  
M  
L  
I  
L  
V  
S  
I  
A  
V  
V  
V  
N  
P  
S  
V  
K  
T  
V  
R  
E  
A  
G  
I  
L  
I  
T  
A  
A  
A  
V  
T  
L  
W  
E  
N  
G  
A  
S  
S  
V  
W  
N  
A  
T  
A  
I  
G  
L  
C  
H  
I  
M  
R  
G  
G  
W  
L  
S  
C  
L  
S  
I  
T  
W  
T  
L  
V  
K  
N  
M  
E  
P  
G  
L  
K  
R  
G  
G  
A  
K  
R  
T  
L  
G  
E  
H  
R  
M  
A  
L  
E  
E  
L  
P  
D  
A  
L  
Q  
T  
I  
A  
L  
I  
A  
L  
S  
V  
M  
T  
G  
V  
F  
L  
L  
M  
Q  
R  
K  
G  
I  
G  
K  
I  
G  
L  
G  
G  
A  
V  
L  
G  
V  
A  
T  
F  
C  
W  
M  
A  
E  
V  
P  
G  
T  
K  
I  
A  
G  
M  
L  
L  
S  
L  
L  
M  
I  
V  
L  
I  
P  
E  
P  
K  
E  
Q  
R  
S  
Q  
T  
D  
N  
Q  
L  
A  
V  
F  
L  
I  
C  
V  
M  
T  
I  
V  
S  
A  
A  
N  
E  
M  
G  
W  
L  
D  
K  
T  
K  
S  
D  
I  
S  
L  
F  
G  
Q  
R  
I  
E  
V  
K  
E  
N  
F  
S  
M  
G  
E  
F  
L  
L  
D  
L  
R  
P  
A  
T  
A  
W  
S  
L  
Y  
A  
V  
T  
A  
V  
L  
T  
P  
L  
L  
K  
H  
L  
I  
T  
S  
D  
Y  
I  
N  
T  
S  
L  
T  
S  
I  
N  
V  
Q  
A  
S  
A  
L  
F  
T  
L  
A  
R  
G  
F  
P  
F  
V  
D  
V  
G  
S  
A  
L  
L  
A  
A  
G  
C  
W  
G  
Q  
V  
T  
L  
T  
V  
T  
A  
A  
T  
L  
L  
F  
C  
H  
Y  
A  
M  
V  
P  
G  
W  
Q  
A  
E  
A  
M  
R  
S  
A  
Q  
R  
R  
T  
A  
G  
I  
M  
K  
N  
A  
V  
V  
D  
G  
I  
V  
A  
T  
D  
V  
P  
E  
L  
R  
T  
P  
I  
M  
Q  
K  
V  
G  
Q  
I  
M  
L  
I  
L  
V  
S  
I  
A  
V  
V  
V  
N  
P  
S  
V  
K  
T  
V  
R  
E  
A  
G  
I  
L  
I  
T  
A  
A  
A  
V  
T  
L  
W  
E  
N  
G  
A  
S  
S  
V  
W  
N  
A  
T  
A  
I  
G  
L  
C  
H  
I  
M  
R  
G  
G  
W  
L  
S  
C  
L  
S  
I  
T  
W  
T  
L  
V  
K  
N  
M  
E  
P  
G  
L  
K  
R  
G  
G  
A  
K  
R  
T  
L  
G  
E  
H  
R  
M  
A  
L  
E  
E  
L  
P  
D  
A  
L  
Q  
T  
I  
A  
L  
I  
A  
L  
S  
V  
M  
T  
G  
V  
F  
L  
L  
M  
Q  
R  
K  
G  
I  
G  
K  
I  
G  
L  
G  
G  
A  
V  
L  
G  
V  
A  
T  
F  
C  
W  
M  
A  
E  
V  
P  
G  
T  
K  
I  
A  
G  
M  
L  
L  
S  
L  
L  
M  
I  
V  
L  
I  
P  
E  
P  
K  
E  
Q  
R  
S  
Q  
T  
D  
N  
Q  
L  
A  
V  
F  
L  
I  
C  
V  
M  
T  
I  
V  
S  
A  
A  
N  
E  
M  
G  
W  
L  
D  
K  
T  
K  
S  
D  
I  
S  
L  
F  
G  
Q  
R  
I  
E  
V  
K  
E  
N  
F  
S  
M  
G  
E  
F  
L  
L  
D  
L  
R  
P  
A  
T  
A  
W  
S  
L  
Y  
A  
V  
T  
A  
V  
L  
T  
P  
L  
L  
K  
H  
L  
I  
T  
S  
D  
Y  
I  
N  
T  
S  
L  
T  
S  
I  
N  
V  
Q  
A  
S  
A  
L  
F  
T  
L  
A  
R  
G  
F  
P  
F  
V  
D  
V  
G  
S  
A  
L  
L  
A  
A  
G  
C  
W  
G  
Q  
V  
T  
L  
T  
V  
T  
A  
A  
T  
L  
L  
F  
C  
H  
Y  
A  
M  
V  
P  
G  
W  
Q  
A  
E  
A  
M  
R  
S  
A  
Q  
R  
R  
T  
A  
G  
I  
M  
K  
N  
A  
V  
V  
D  
G  
I  
V  
A  
T  
D  
V  
P  
E  
L  
R  
T  
P  
I  
M  
Q  
K  
V  
G  
Q  
I  
M  
L  
I  
L  
V  
S  
I  
A  
V  
V  
V  
N  
P  
S  
V  
K  
T  
V  
R  
E  
A  
G  
I  
L  
I  
T  
A  
A  
A  
V  
T  
L  
W  
E  
N  
G  
A  
S  
S  
V  
W  
N  
A  
T  
A  
I  
G  
L  
C  
H  
I  
M  
R  
G  
G  
W  
L  
S  
C  
L  
S  
I  
T  
W  
T  
L  
V  
K  
N  
M  
E  
P  
G  
L  
K  
R  
G  
G  
A  
K  
R  
T  
L  
G  
E  
H  
R  
M  
A  
L  
E  
E  
L  
P  
D  
A  
L  
Q  
T  
I  
A  
L  
I  
A  
L  
S  
V  
M  
T  
G  
V  
F  
L  
L  
M  
Q  
R  
K  
G  
I  
G  
K  
I  
G  
L  
G  
G  
A  
V  
L  
G  
V  
A  
T  
F  
C  
W  
M  
A  
E  
V  
P  
G  
T  
K  
I  
A  
G  
M  
L  
L  
S  
L  
L  
M  
I  
V  
L  
I  
P  
E  
P  
K  
E  
Q  
R  
S  
Q  
T  
D  
N  
Q  
L  
A  
V  
F  
L  
I  
C  
V  
M  
T  
I  
V  
S  
A  
A  
N  
E  
M  
G  
W  
L  
D  
K  
T  
K  
S  
D  
I  
S  
L  
F  
G  
Q  
R  
I  
E  
V  
K  
E  
N  
F  
S  
M  
G  
E  
F  
L  
L  
D  
L  
R  
P  
A  
T  
A  
W  
S  
L  
Y  
A  
V  
T  
A  
V  
L  
T  
P  
L  
L  
K  
H  
L  
I  
T  
S  
D  
Y  
I  
N  
T  
S  
L  
T  
S  
I  
N  
V  
Q  
A  
S  
A  
L  
F  
T  
L  
A  
R  
G  
F  
P  
F  
V  
D  
V  
G  
S  
A  
L  
L  
A  
A  
G  
C  
W  
G  
Q  
V  
T  
L  
T  
V  
T  
A  
A  
T  
L  
L  
F  
C  
H  
Y  
A  
M  
V  
P  
G  
W  
Q  
A  
E  
A  
M  
R  
S  
A  
Q  
R  
R  
T  
A  
G  
I  
M  
K  
N  
A  
V  
V  
D  
G  
I  
V  
A  
T  
D  
V  
P  
E  
L  
R  
T  
P  
I  
M  
Q  
K  
V  
G  
Q  
I  
M  
L  
I  
L  
V  
S  
I  
A  
V  
V  
V  
N  
P  
S  
V  
K  
T  
V  
R  
E  
A  
G  
I  
L  
I  
T  
A  
A  
A  
V  
T  
L  
W  
E  
N  
G  
A  
S  
S  
V  
W  
N  
A  
T  
A  
I  
G  
L  
C  
H  
I  
M  
R  
G  
G  
W  
L  
S  
C  
L  
S  
I  
T  
W  
T  
L  
V  
K  
N  
M  
E  
P  
G  
L  
K  
R  
G  
G  
A  
K  
R  
T  
L  
G  
E  
H  
R  
M  
A  
L  
E  
E  
L  
P  
D  
A  
L  
Q  
T  
I  
A  
L  
I  
A  
L  
S  
V  
M  
T  
G  
V  
F  
L  
L  
M  
Q  
R  
K  
G  
I  
G  
K  
I  
G  
L  
G  
G  
A  
V  
L  
G  
V  
A  
T  
F  
C  
W  
M  
A  
E  
V  
P  
G  
T  
K  
I  
A  
G  
M  
L  
L  
S  
L  
L  
M  
I  
V  
L  
I  
P  
E  
P  
K  
E  
Q  
R  
S  
Q  
T  
D  
N  
Q  
L  
A  
V  
F  
L  
I  
C  
V  
M  
T  
I  
V  
S  
A  
A  
N  
E  
M  
G  
W  
L  
D  
K  
T  
K  
S  
D  
I  
S  
L  
F  
G  
Q  
R  
I  
E  
V  
K  
E  
N  
F  
S  
M  
G  
E  
F  
L  
L  
D  
L  
R  
P  
A  
T  
A  
W  
S  
L  
Y  
A  
V  
T  
A  
V  
L  
T  
P  
L  
L  
K  
H  
L  
I  
T  
S  
D  
Y  
I  
N  
T  
S  
L  
T  
S  
I  
N  
V  
Q  
A  
S  
A  
L  
F  
T  
L  
A  
R  
G  
F  
P  
F  
V  
D  
V  
G  
S  
A  
L  
L  
A  
A  
G  
C  
W  
G  
Q  
V  
T  
L  
T  
V  
T  
A  
A  
T  
L  
L  
F  
C  
H  
Y  
A  
M  
V  
P  
G  
W  
Q  
A  
E  
A  
M  
R  
S  
A  
Q  
R  
R  
T  
A  
G  
I  
M  
K  
N  
A  
V  
V  
D  
G  
I  
V  
A  
T  
D  
V  
P  
E  
L  
R  
T  
P  
I  
M  
Q  
K  
V  
G  
Q  
I  
M  
L  
I  
L  
V  
S  
I  
A  
V  
V  
V  
N  
P  
S  
V  
K  
T  
V  
R  
E  
A  
G  
I  
L  
I  
T  
A  
A  
A  
V  
T  
L  
W  
E  
N  
G  
A  
S  
S  
V  
W  
N  
A  
T  
A  
I  
G  
L  
C  
H  
I  
M  
R  
G  
G  
W  
L  
S  
C  
L  
S  
I  
T  
W  
T  
L  
V  
K  
N  
M  
E  
P  
G  
L  
K  
R  
G  
G  
A  
K  
R  
T  
L  
G  
E  
H  
R  
M  
A  
L  
E  
E  
L  
P  
D  
A  
L  
Q  
T  
I  
A  
L  
I  
A  
L  
S  
V  
M  
T  
G  
V  
F  
L  
L  
M  
Q  
R  
K  
G  
I  
G  
K  
I  
G  
L  
G  
G  
A  
V  
L  
G  
V  
A  
T  
F  
C  
W  
M  
A  
E  
V  
P  
G  
T  
K  
I  
A  
G  
M  
L  
L  
S  
L  
L  
M  
I  
V  
L  
I  
P  
E  
P  
K  
E  
Q  
R  
S  
Q  
T  
D  
N  
Q  
L  
A  
V  
F  
L  
I  
C  
V  
M  
T  
I  
V  
S  
A  
A  
N  
E  
M  
G  
W  
L  
D  
K  
T  
K  
S  
D  
I  
S  
L  
F  
G  
Q  
R  
I  
E  
V  
K  
E  
N  
F  
S  
M  
G  
E  
F  
L  
L  
D  
L  
R  
P  
A  
T  
A  
W  
S  
L  
Y  
A  
V  
T  
A  
V  
L  
T  
P  
L  
L  
K  
H  
L  
I  
T  
S  
D  
Y  
I  
N  
T  
S  
L  
T  
S  
I  
N  
V  
Q  
A  
S  
A  
L  
F  
T  
L  
A  
R  
G  
F  
P  
F  
V  
D  
V  
G  
S  
A  
L  
L  
A  
A  
G  
C  
W  
G  
Q  
V  
T  
L  
T  
V  
T  
A  
A  
T  
L  
L  
F  
C  
H  
Y  
A  
M  
V  
P  
G  
W  
Q  
A  
E  
A  
M  
R  
S  
A  
Q  
R  
R  
T  
A  
G  
I  
M  
K  
N  
A  
V  
V  
D  
G  
I  
V  
A  
T  
D  
V  
P  
E  
L  
R  
T  
P  
I  
M  
Q  
K  
V  
G  
Q  
I  
M  
L  
I  
L  
V  
S  
I  
A  
V  
V  
V  
N  
P  
S  
V  
K  
T  
V  
R  
E  
A  
G  
I  
L  
I  
T  
A  
A  
A  
V  
T  
L  
W  
E  
N  
G  
A  
S  
S  
V  
W  
N  
A  
T  
A  
I  
G  
L  
C  
H  
I  
M  
R  
G  
G  
W  
L  
S  
C  
L  
S  
I  
T  
W  
T  
L  
V  
K  
N  
M  
E  
P  
G  
L  
K  
R  
G  
G  
A  
K  
R  
T  
L  
G  
E  
H  
R  
M  
A  
L  
E  
E  
L  
P  
D  
A  
L  
Q  
T  
I  
A  
L  
I  
A  
L  
S  
V  
M  
T  
G  
V  
F  
L  
L  
M  
Q  
R  
K  
G  
I  
G  
K  
I  
G  
L  
G  
G  
A  
V  
L  
G  
V  
A  
T  
F  
C  
W  
M  
A  
E  
V  
P  
G  
T  
K  
I  
A  
G  
M  
L  
L  
S  
L  
L  
M  
I  
V  
L  
I  
P  
E  
P  
K  
E  
Q  
R  
S  
Q  
T  
D  
N  
Q  
L  
A  
V  
F  
L  
I  
C  
V  
M  
T  
I  
V  
S  
A  
A  
N  
E  
M  
G  
W  
L  
D  
K  
T  
K  
S  
D  
I  
S  
L  
F  
G  
Q  
R  
I  
E  
V  
K  
E  
N  
F  
S  
M  
G  
E  
F  
L  
L  
D  
L  
R  
P  
A  
T  
A  
W  
S  
L  
Y  
A  
V  
T  
A  
V  
L  
T  
P  
L  
L  
K  
H  
L  
I  
T  
S  
D  
Y  
I  
N  
T  
S  
L  
T  
S  
I  
N  
V  
Q  
A  
S  
A  
L  
F  
T  
L  
A  
R  
G  
F  
P  
F  
V  
D  
V  
G  
S  
A  
L  
L  
A  
A  
G  
C  
W  
G  
Q  
V  
T  
L  
T  
V  
T  
A  
A  
T  
L  
L  
F  
C  
H  
Y  
A  
M  
V  
P  
G  
W  
Q  
A  
E  
A  
M  
R  
S  
A  
Q  
R  
R  
T  
A  
G  
I  
M  
K  
N  
A  
V  
V  
D  
G  
I  
V  
A  
T  
D  
V  
P  
E  
L  
R  
T  
P  
I  
M  
Q  
K  
V  
G  
Q  
I  
M  
L  
I  
L  
V  
S  
I  
A  
V  
V  
V  
N  
P  
S  
V  
K  
T  
V  
R  
E  
A  
G  
I  
L  
I  
T  
A  
A  
A  
V  
T  
L  
W  
E  
N  
G  
A  
S  
S  
V  
W  
N  
A  
T  
A  
I  
G  
L  
C  
H  
I  
M  
R  
G  
G  
W  
L  
S  
C  
L  
S  
I  
T  
W  
T  
L  
V  
K  
N  
M  
E  
P  
G  
L  
K  
R  
G  
G  
A  
K  
R  
T  
L  
G  
E  
H  
R  
M  
A  
L  
E  
E  
L  
P  
D  
A  
L  
Q  
T  
I  
A  
L  
I  
A  
L  
S  
V  
M  
T  
G  
V  
F  
L  
L  
M  
Q  
R  
K  
G  
I  
G  
K  
I  
G  
L  
G  
G  
A  
V  
L  
G  
V  
A  
T  
F  
C  
W  
M  
A  
E  
V  
P  
G  
T  
K  
I  
A  
G  
M  
L  
L  
S  
L  
L  
M  
I  
V  
L  
I  
P  
E  
P  
K  
E  
Q  
R  
S  
Q  
T  
D  
N  
Q  
L  
A  
V  
F  
L  
I  
C  
V  
M  
T  
I  
V  
S  
A  
A  
N  
E  
M  
G  
W  
L  
D  
K  
T  
K  
S  
D  
I  
S  
L  
F  
G  
Q  
R  
I  
E  
V  
K  
E  
N  
F  
S  
M  
G  
E  
F  
L  
L  
D  
L  
R  
P  
A  
T  
A  
W  
S  
L  
Y  
A  
V  
T  
A  
V  
L  
T  
P  
L  
L  
K  
H  
L  
I  
T  
S  
D  
Y  
I  
N  
T  
S  
L  
T  
S  
I  
N  
V  
Q  
A  
S  
A  
L  
F  
T  
L  
A  
R  
G  
F  
P  
F  
V  
D  
V  
G  
S  
A  
L  
L  
A  
A  
G  
C  
W  
G  
Q  
V  
T  
L  
T  
V  
T

mat\_peptide 2470. .3525  
/product="non-structural protein 1"  
/note="NS1"  
mat\_peptide 3526. .4218  
/product="non-structural protein 2A"  
/note="NS2A"  
mat\_peptide 4219. .4611  
/product="non-structural protein 2B"  
/note="NS2B"  
mat\_peptide 4612. .6468  
/product="non-structural protein 3"  
/note="NS3"  
mat\_peptide 6469. .6915  
/product="non-structural protein 4A"  
/note="NS4A"  
mat\_peptide 6916. .7680  
/product="non-structural protein 4B"  
/note="NS4B"  
mat\_peptide 7681. .10395  
/product="non-structural protein 5"  
/note="NS5"

## ORIGIN

Db	1376	ACGAAGTGGCCATTTGTCCATGGACCAACTACTGTGGAGTCGCACGGAAACTACCCCA	1435
Qy	481	CACAGATAAGGGCTACCCAAGCAGGAAGGTCAGCATAACTCCATCGGCACCACCTACA	540
Db	1436	CACAGGTTGGAGCCACTCAGGCAGGGAGATTCAAGCATCACTCCTGC GGCGCCTTCATACA	1495
Qy	541	CGCTGAAGTGGGTGAGTATGGTGAGGTACAGTTGACTGTGAGGCCACGGTCAGGAATAG	600
Db	1496	CACTAAAGCTTGGAGAATATGGAGAGGTGACAGTGGACTGTGAACCACGGTCAGGGATTG	1555
Qy	601	ACACTAGCGCTTACTACGTTATGTCAGTGGGTGCGAAGTCCTCTGGTTCACCGAGAAT	660
Db	1556	ACACCAATGCATACTACGTGATGACTGTTGGAACAAAGACGTTCTGGTCCATCGTGAGT	1615
Qy	661	GGTTTATGGACCTGAACCTTCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG	720
Db	1616	GGTTCATGGATCTAACCTCCCTGGAGCAGTGCTGGAAGTACTGTGTGGAGGAACAGAG	1675
Qy	721	AAACACTGATGGAGTTGAAGAACCTCATGCCACCAAACAATCTGCGTAGCTCTAGGGT	780
Db	1676	AGACGTTAACGGAGTTGAGGAACCACACGCCACGAAGCAGTCTGTGATAGCATTGGGCT	1735
Qy	781	CGCAGGAAGGTGCCTTGCACCAAGCTCTGGCTGGAGCAATTCTGTTGAGTTCTCAAGCA	840
Db	1736	CACAAGAGGGAGCTGCATCAAGCTTGGCTGGAGCCATTCTGTGGAATTTCAGCAAGCA	1795
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTAAGATGGAGAAGTTGCAGC	900
Db	1796	ACACTGTCAAGTTGACGTCGGGTCAATTGAAGTGTAGAGTGAAGATGGAAAAATTGCAGT	1855
Qy	901	TGAAGGGAACACATATGGTGTATGCTCAAAGCATTCAAATTGCTAGGACTCCGCTG	960
Db	1856	TGAAGGGAACACCTATGGCGTCTGTTCAAAGGCTTCAAGTTCTGGGACTCCGCGAG	1915
Qy	961	ACACTGGTCATGGAACGGTGGTGTGGAACTGCAGTATACCGGAAAAGACGGCCTTGCA	1020
Db	1916	ACACAGGTACGGCACTGTGGTGTGGAAATTGCAAGTACACTGGCACGGATGGACCTTGCA	1975
Qy	1021	AAGTGCCCATTTCTTGTGGCTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGA	1080
Db	1976	AAGTTCCATCTCGTCAGTGGCTTCATTGAACGACCTAACGCCAGTGGCAGATTGGTCA	2035
Qy	1081	CTGTGAATCCATTGTGTCTGTGGCTACGGCCAACCTCGAAGGTTTGATTGAACCGAAC	1140
Db	2036	CTGTCAACCCCTTTGTTCAAGTGGCCACGGCCAACGCTAAGGTCTGATTGAATTGGAAC	2095
Qy	1141	CCCCGTTAGTGAECTTACATCGTGGTGGGGAGAGGAGAACAGCAGATAAACCAACT	1200
Db	2096	CACCCCTTGGAGACTCATACATAGTGGTGGCAGAGGAGAACACAGATCAATCACCATT	2155
Qy	1201	GGCACAAATCTGGGAGCAGTATTGGAAAGGCTTCACCACTACACTCAGAGGAGCTAAC	1260
Db	2156	GGCACAAAGTCTGGAAGCAGCATTGGCAAAGCCTTACAACCACCTCAAAGGAGCGCAGA	2215
Qy	1261	GACTTGCAGCTCTTGGAGACACTGCCTGGGATTTGGATCAGTGGAGGGTTTCACCT	1320
Db	2216	GACTAGCCGCTCTAGGAGACACAGCTTGGGACTTGGATCAGTGGAGGGTGTACCT	2275

Qy	1321 CGGTAGGGAAAGCCATACACCAAGTTTGGAGGAGCCTTAGATCACTCTTGGAGGGA 1380 
Db	2276 CAGTTGGGAAGGCTGTCCATCAAGTGTTCGGAGGAGCATTCCGCTCACTGTTGGAGGCA 2335
Qy	1381 TGTCTGGATCACACAGGGCTCTGGAGCTCTGCTGTGGATGGAAATTAACGCC 1440 
Db	2336 TGTCTGGATAACGCAAGGATTGCTGGGGCTCTCCTGTTGGATGGCATTAAATGCTC 2395
Qy	1441 GTGACAGGTCAATTGCTATGACGTTCTGCAGTTGGAGGAGTCTGCTCTCCTTCGG 1500 
Db	2396 GTGATAGGTCCATAGCTCTCACGTTCTGCAGTTGGAGGAGTCTGCTCTCCTCTCCG 2455
Qy	1501 TCAACGTCCATGCTG 1515 
Db	2456 TGAACGTGCATGCTG 2470

RESULT 13

DQ164190

LOCUS DQ164190 11029 bp RNA linear VRL 18-NOV-2005  
 DEFINITION West Nile virus isolate NY 2003 Suffolk, complete genome.  
 ACCESSION DQ164190  
 VERSION DQ164190.1 GI:76781539  
 KEYWORDS .  
 SOURCE West Nile virus (WNV)  
 ORGANISM West Nile virus  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 Flavivirus; Japanese encephalitis virus group.  
 REFERENCE 1 (bases 1 to 11029)  
 AUTHORS Davis,C.T., Ebel,G.D., Lanciotti,R.S., Brault,A.C., Guzman,H.,  
 Siirin,M., Lambert,A., Parsons,R.E., Beasley,D.W., Novak,R.J.,  
 Elizondo-Quiroga,D., Green,E.N., Young,D.S., Stark,L.M.,  
 Drexler,M.A., Artsob,H., Tesh,R.B., Kramer,L.D. and Barrett,A.D.  
 TITLE Phylogenetic analysis of North American West Nile virus isolates,  
 2001-2004: Evidence for the emergence of a dominant genotype  
 JOURNAL Virology 342 (2), 252-265 (2005)  
 PUBMED 16137736  
 REFERENCE 2 (bases 1 to 11029)  
 AUTHORS Davis,T.C., Ebel,G.D., Lanciotti,R.S. and Brault,A.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-AUG-2005) Pathology, University of Texas Medical  
 Branch, 301 University Blvd., Galveston, TX 77550, USA  
 FEATURES Location/Qualifiers  
 source 1. .11029  
 /organism="West Nile virus"  
 /mol\_type="genomic RNA"  
 /isolate="NY 2003 Suffolk"  
 /specific\_host="American crow"  
 /db\_xref="taxon:11082"  
 /country="USA: NY, Suffolk"  
 /collection\_date="2003"  
 CDS 97. .10398  
 /note="encodes C, prM, E, NS1, NS2A, NS2B, NS3, NS4A,  
 NS4B, NS5"  
 /codon\_start=1  
 /product="polyprotein precursor"

/protein\_id="ABA54579.1"  
/db\_xref="GI: 76781540"  
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSIGLKRAMLSLIDKGKPI  
RFVLALLAFFFRTAIAPTRAVLNWRGVNKQTAMKHLLSFKKEGLTTSAINRRSSKQ  
KKRGGKTGIAVMIGLIASVGAVTLSNFQGVMMTVNATDVTDVITIPTAAGKNLCIVR  
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAYVRYGRCTKTRHSRRSRSLT  
VQTHGESTLANKGAWMDSTKATRLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV.  
VFVVLVLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDCVTIMSKDKPTIDVK  
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG  
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQAG  
ATQAGRFSITPAAPSYTLLGEYGEVTVDCEPRSGIDTNAYYMTVGTKTFLVHREWF  
MDILNPWSSAGSTVWRNRETLMEFEETHATKQSVIALGSQEGALHQALAGAIPEFSS  
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSSKAFKFLGTPADTGHGTVVLELQYTGTDG  
PCKVPISSVASLNDLTPVGRLVTFVNPFSVATANAKVIELEPPFGDSYIVVGRGEQQ  
INHHWHKSGSSIGKAFTTLKGAAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF  
RSLFGGMSWITQGLLGALLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID  
ISRQEELRCGSGVFIHNDEAWMDRYKYPETPQGLAKIIQKAHEGVCGRLRSVRLEH  
QMWEAVKDELNTLLKENGVDLSVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSL  
FAPELANNTFVVDGPETKECPTQNRAWSLEVEDFGFGLTSTRMFLKVRESNTTECDS  
KIIGTAVKNLAIHSDSLSDYWIERSLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES  
DLIIPVTLAGPRSNNRPGYKTQNQGPWDEGRVEIDFDYCPGTVTLSESCGRHPA  
TRTTTESGKLITDWCCRSC TLPLR YQTD SG C WY G M E I R P Q R H D E K T L V Q S Q V N A Y N A  
DMIDPFQLGLLVVFLATQEVLRKRWTAKISM P A I L I A L L V L V F G G I T Y T D V L R Y V I L V  
GAFAAESNSGGDVVH L A L M A T F K I Q P V F M V A S F L K A R W T N Q E N I I L M L A A V F F Q M A Y H  
DARQILLWEIPDVNLNSLAVAWMILRAITFTTSN V V V P L L A L L T P G L R C L N L D V Y R I L  
LLMVGIGSLIREKRSAAKKKGASLLCLALASTGLFNPMLAAGLIACDPNRKRGWPA  
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW  
ESDAEITGS SERV D VR L D D G N F Q L M N D P G A P W K I W M L R M V C L A I S A Y T P W A I L P S V V  
GFWITLQYT KRGGV L W D T P S P K E Y K K G D T T G V Y R I M T R G L L G S Y Q A G A G V M V E G V F H  
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMI V V E P G K N  
VKVQTKPGVFKTPEGEIGAVTLDFTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS  
AIVQGERMDEPI PAGFEP EMLRKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV  
LAPTRVVAEAMA E ALRGLPIRYQTS A V P R E H N G N E I V D V M C H A T L T H R L M S P H R V P N Y  
NL F V M D E A H F T D P A S I A A R G Y I S T K V E L G E A A A I F M T A T P P G T S D P F P E S N S P I S D L Q  
TEI P DRAWN S G Y E W I T E Y T G K T V W F V P S V K M G N E I A L C L Q R A G K V V Q L N R K S Y E T E Y  
PKCKNDDWD F V I T T D I S E M G A N F K A S R V I D S R K S V K P T I I T E G E G R V I L G E P S A V T A A  
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHTEARIMLDNINMPNGLIAQFYQ  
PEREKVYTMGEYRLRGEERKNFLELLRTADLPVWLAYKVAAGVSYHDRRCFDGPR  
TNTILEDNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFA SGKRSQIGLIEV  
LGKMP EHFMGKTWEALDTM Y V V A T A E K G G R A H R M A L E E L P D A L Q T I A L I A L S V M T M G  
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGM L L S L L M I V L I P E P E K  
QRSQTDNQLAVFLICVMTLVS A V A A N E M G W L D K T K S D I S S L F G Q R I E V K E N F S M G E F L  
LDLRPATAWSLYAVTTAVLTPLLKHLITS D Y I N T S L T S I N V Q A S A L F T L A R G F P F V D V  
GVSALLAAGCWGQVLT V T A T L L F C H Y A Y M V P G W Q A E A M R S A Q R R T A A G I M K N A  
VVDGIVATDVP E L E R T P I M Q K V G Q I M L I L V S L A A V V N P S V K T V R E A G I L I T A A V  
TLWENGASSVWNATT A I G L C H I M R G G W L S C L S I T W T L I K N M E K P G L K R G G A K G R T L G E  
VWKERLNQMTKEE FTRYRKEAII E V D R S A A K H A R K E G N V T G G H P V S R G T A K L R W L V E R  
RFL EPVGKVIDLGCRGGWCYYMATQKRVQEVRYT KGGPGHEEPQLVQSYGWNIVTM  
KSGVDFYRPS E C C D T L L C D I G E S S S A E V E E H R T I R V L E M V E D W L H R G P R E F C V K V L  
CPYMPKVI E K M E L L Q R R Y G G G L V R N P L S R N S T H E M Y W V S R A S G N V V H S V N M T S Q V L L G  
RMEKRTWKGPOQEEDVN L G S G T R A V G K P L L N S D T R K I K N R I E R L R E Y S S T W H H D E N H  
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWD T I T N V T T M A M T D T T P F G Q Q R V F K  
EKVDTKAPEPPEGV KYVLNETTNWLWAFLAREKRPRMC SREEFIRKVN S N A A L G A M F E  
EQNQWRSAREA E D P K F W E M V D E E R E A H L R G E C H T C I Y N M M G K R E K K P G E F G K A K G S R  
AIWFMWL GARFLEFEALGFLNEDHWLGRKN S G G G V E G L Q K L G Y I L R E V G T R P G G K I  
YADDTAGWDTRITRADLENEAKVLELLDGEHRR LARAI I E L T Y R H K V V K M R P A A D G R  
TVM DVISREDQRGSGQVVTYALNTFTNLA V Q L V R M M E G E G V I G P D D V E K L T K G K G P K V

RTWLFENGEERLSRMAVSGDDCVVKPLDDRATSLHFLNAMSKVRKDIQEWPSTGKY  
DWQQVPFCSNHFTELIMKDGRTLVPCRGQDELVGRARI SPAGWNVRDTACLAKSYA  
QMWLLLYFHRRDLRLMANAICSAVPVNWPTGRTTWSIHAGGEWTTEDMLEVWNRVW  
IEENEWMEDKTPVEKWSDVYPYGKREDIWCGLIGTRARATWAENIQVAINQVRAIIG  
DEKYVDYMSLKRUYEDTTLVEDTVL"

## ORIGIN

Query Match 65.7%; Score 998.2; DB 10; Length 11029;  
 Best Local Similarity 78.7%; Pred. No. 0;  
 Matches 1192; Conservative 0; Mismatches 323; Indels 0; Gaps 0;

Qy	1	CGGAATTCTAGCTTCAACTGTTAGGAATGAGCAACAGGGACTCCTGGAGGGAGTGTCTG	60
Db	956	CAGCTTACAGCTTCAACTGCCTTGGAAATGAGCAACAGAGACTTCTTGGAAAGGAGTGTCTG	1015
Qy	61	GAGCTACATGGGTTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAAG	120
Db	1016	GAGCAACATGGGTGGATTGTTCTCGAAGGCGACAGCTGCGTGACTATCATGTCTAAGG	1075
Qy	121	ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAACATCGCAGATGTGC	180
Db	1076	ACAAGCCTACCATCGATGTGAAGATGATGAATATGGAGGCGCCAAACCTGGCAGAGGTCC	1135
Qy	181	GTAGCTACTGCTACTTAGCTCGGTAGTCAGTGATCTGTCAACAAAAGCCCGTGTCCAACCA	240
Db	1136	GCAGTTATTGCTATTGGCTACCGTCAGCGATCTCCACCAAAGCTGCGTCCCCGACCA	1195
Qy	241	TGGGTGAAGCTCACAAACGAGAAAAGAGCCGACCCCTGCCTTGTGCAAGCAAGGCGTCG	300
Db	1196	TGGGAGAAGCTCACAAACGTGCTGACCCAGCTTGTGCAAGGAGTGG	1255
Qy	301	TAGACAGAGGATGGGGAATGGATGCGACTGTTGGAAAGGGGAGCATTGACACATGTG	360
Db	1256	TGGACAGGGGCTGGGCAACGGCTGCGGACTATTGGCAAGGAAGCATTGACACATGCG	1315
Qy	361	CAAAGTTGCCTGTACAACCAAGGCAACTGGTTGATTATCCAGAAGGAAAACATCAAGT	420
Db	1316	CCAAATTGCTGCTCTACCAAGGCAATAGGAAGAACATCTGAAAGAGAAATATCAAGT	1375
Qy	421	ACGAGGTTGCCATATTGTGCATGGCCGACGACTGTCGAATCACATGGCAATTATTCAA	480
Db	1376	ACGAAGTGGCCATTGGTCCATGGACCAACTACTGTGGAGTCGCACGGAAACTACTCCA	1435
Qy	481	CACAGATAGGGCTACCCAAAGCAGGAAGGTCAGCATAACTCCATCGGCACCACCTACA	540
Db	1436	CACAGGCTGGAGCCACTCAGGCAGGGAGATTCAAGCATCACTCCTGCGGCGCTTCATACA	1495
Qy	541	CGCTGAAGTTGGGTGAGTATGGTGGAGTCACAGTGACTGTGAGGCCACGGTCAGGAATAG	600
Db	1496	CACTAAAGCTGGAGAATATGGAGAGGTGACAGTGGACTGTGAACCACGGTCAGGGATTG	1555
Qy	601	ACACTAGCGCTTACTACGTTATGTCAGTGGGTGCGAACGTCCTCTGGTCACCGAGAAT	660
Db	1556	ACACCAATGCATACTACGTGATGACTGTTGAAACAAAGACGTTCTGGTCCATCGTGAGT	1615
Qy	661	GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAAGCACAACGTGGAGGAACCGGG	720

Db	1616	GGTTCATGGACCTAACCTCCCTGGAGCAGTGCTGGAAGTACTGTGTGGAGGAACAGAG	1675
Qy	721	AAACACTGATGGAGTTGAAGAACCTCATGCCACCAACAATCTGTCGTAGCTCTAGGGT	780
Db	1676	AGACGTTAATGGAGTTGAGGAACCACACGCCACGAAGCAGTCTGTGATAGCATTGGGCT	1735
Qy	781	CGCAGGAAGGTGCCTTGACCAAGCTCTGGCTGGAGCAATTCTGTTGAGTCTCAAGCA	840
Db	1736	CACAAGAGGGAGCTCTGCATCAAGCTTGGCTGGAGCCATTCTGTGGAATTCTCAAGCA	1795
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC	900
Db	1796	ACACTGTCAAGTTGACGTCGGTCATTGAAGTGTAGAGTGAAGATGGAAAAATTGCAGT	1855
Qy	901	TGAAGGGAACAACATATGGTGTATGCTAAAAGCATTCAAATTGCTAGGACTCCGCTG	960
Db	1856	TGAAGGGAACAACCTATGGCGTCTGTTCAAAGGTTCAAGTTCTGGACTCCGCGAG	1915
Qy	961	ACACTGGTCATGGAACGGTGGTGGAACTGCAGTATAACGGAAAAGACGGGCCTTGCA	1020
Db	1916	ACACAGGTACGGCACTGTGGTGGATTGAGTACACTGGCACGGATGGACCTTGCA	1975
Qy	1021	AAGTGCCCATTCTTCTGTGGCTCCCTGAACGACCTACACCCGTTGGAAGGCTGGTGA	1080
Db	1976	AAGTTCCATCTCGTCAGTGGCTTCATTGAACGACCTAACGCCAGTGGCAGATTGGTCA	2035
Qy	1081	CTGTGAATCCATTGTGTCTGTGGCTACGGCCAACTCGAAGGTTTGATTGAACCGAAC	1140
Db	2036	CTGTCAACCCTTTGTTCACTGGCCACGGCAACGCTAAGGTCTGATTGAATTGGAAC	2095
Qy	1141	CCCCGTTAGTGAECTTACATCGTGGTGGGAGAGGAGAACAGCAGATAAACCAACT	1200
Db	2096	CACCCCTTGGAGACTCATACATAGTGGTGGCAGAGGAGAACACAGATCAATCACCATT	2155
Qy	1201	GGCACAAATCTGGAGCAGTATTGAAAGGCTTCACCACTACACTCAGAGGAGCTAAC	1260
Db	2156	GGCACAAAGTCTGGAAGCAGCATTGGCAAAGCCTTACAACCACCTCAAAGGAGCGCAGA	2215
Qy	1261	GACTTGCAGCTCTGGAGACACTGCCTGGGATTTGGATCAGTCGGAGGGGTTTCACCT	1320
Db	2216	GACTAGCCGCTCTAGGAGACACAGCTGGACTTGGATCAGTTGGAGGGGTGTTCACCT	2275
Qy	1321	CGGTAGGGAAAGCCATACACCAAGTTTGAGGAGCCTTAGATCACTCTTGGAGGGA	1380
Db	2276	CAGTTGGGAAGGCTGTCATCAAGTGGTGGAGGAGCATTCCGCTCACTGTTGGAGGCA	2335
Qy	1381	TGTCCTGGATCACACAGGGCTTCTGGAGCTTTCTGCTGTGGATGGAAATTACGCC	1440
Db	2336	TGTCCTGGATAACGCAAGGATTGCTGGGGCTCTCCTGTTGTGGATGGCATCAATGCTC	2395
Qy	1441	GTGACAGGTCAATTGCTATGACGTTCTGGCGGTTGGAGGAGTCTTGCTCTCCTTCGG	1500
Db	2396	GTGATAGGTCCATAGCTCTCACGTTCTCGCAGTTGGAGGAGTCTGCTCTCCTCTCG	2455
Qy	1501	TCAACGTCCATGCTG	1515
Db	2456	TGAACGTGCATGCTG	2470

RESULT 14

AY371271

LOCUS AY371271 2004 bp RNA linear VRL 25-NOV-2003  
DEFINITION West Nile virus strain TM171-03 polyprotein gene, partial cds.  
ACCESSION AY371271  
VERSION AY371271.1 GI:38224786  
KEYWORDS .  
SOURCE West Nile virus (WNV)  
ORGANISM West Nile virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
Flavivirus; Japanese encephalitis virus group.  
REFERENCE 1 (bases 1 to 2004)  
AUTHORS Estrada-Franco,J.G., Navarro-Lopez,R., Beasley,D.W.C., Coffey,L.,  
Carrara,A.-S., Travassos da Rosa,A., Clements,T., Wang,E.,  
Ludwig,G.V., Campomanes Cortes,A., Paz Ramirez,P., Tesh,R.B.,  
Barrett,A.D.T. and Weaver,S.C.  
TITLE West Nile virus in Mexico: evidence of widespread circulation since  
July, 2002  
JOURNAL Emerging Infect. Dis. 9 (12), 1604-1607 (2003)  
REFERENCE 2 (bases 1 to 2004)  
AUTHORS Beasley,D.W.C., Estrada-Franco,J.G., Tesh,R.B., Weaver,S.C. and  
Barrett,A.D.T.  
TITLE Direct Submission  
JOURNAL Submitted (20-AUG-2003) Pathology, University of Texas Medical  
Branch, 301 University Blvd., Galveston, TX 77555-0609, USA  
FEATURES Location/Qualifiers  
source 1. .2004  
/organism="West Nile virus"  
/mol\_type="genomic RNA"  
/strain="TM171-03"  
/db\_xref="taxon:11082"  
/country="Mexico"  
CDS <1. .>2004  
/codon\_start=1  
/product="polyprotein"  
/protein\_id="AAR14153.1"  
/db\_xref="GI:38224787"  
/translation="VTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVRAMDVGYM  
CDDTITYECPVLSAGNDPEDI DCWCTKSAVVYVRYGRCTKTRHSRRSRSRSLTVQTHGES  
TLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVTGWLGSNTMQRVVFVLLL  
LVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMS KDKPTIDVKMMNMEA  
NLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWGNCGGLFG  
KGSIIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYPTQVGATQAGRF  
SITPAAPSYT LKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWFM DLNLPW  
SSAGSTVWRNRETLMEFEEPHATKQS VIALGSQEGALHQALAGAI PVEFSSNTVKLTS  
GHLKCRVKMEKLQLKGTTYGVC SKAFKFLGPADTGHGTVVLELQYTGTGDPCKVPIS  
SVASLNDLTPVGR LTVNP FV SVATANAKV LIELEPPFGDSYIVVGRGEQQINHHWHK  
SGSSIGKAFTTTLKG AQR L A ALGDTAWDFGSVGGVFTSVG KAVHQVFGGAFRSLF GGM  
SWITQGLLGALLWMGINARD RSI ALTFLAVGGVLLFLSVNVHA"  
mat\_peptide 1. .276  
/product="pre-membrane protein prM"  
mat\_peptide 277. .501  
/product="membrane protein M"  
mat\_peptide 502. .2004  
/product="envelope protein E"

## ORIGIN

Query Match 65.6%; Score 997.2; DB 10; Length 2004;  
Best Local Similarity 78.7%; Pred. No. 0;  
Matches 1191; Conservative 0; Mismatches 323; Indels 0; Gaps 0;

Qy 1 CGGAATTCACTGTTAGGAATGAGCAACAGGGACTTCTGGAGGGAGTGTCTG 60  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 491 CAGCTTACAGCTCACTGCCTTGAATGAGCAACAGAGACTTCTTGAAGGAGTGTCTG 550

Qy 61 GAGCTACATGGGTTGATCTGGTACTGGAAGGAGACAGTTGTGACCATAATGTCAAAAG 120  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 551 GAGCAACATGGGTGGATTGGTCTCGAAGGCAGCTGCGTACTATCATGTCTAAGG 610

Qy 121 ACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCTAACATCGCAGATGTGC 180  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 611 ACAAGCCTACCATCGATGTGAAGATGATGAATATGGAGGCGAACCTGGCAGAGGTCC 670

Qy 181 GTAGCTACTGCTACTTAGCTCGTCAGTGATCTGTCAACAAAAGCCCGTGTCCAACCA 240  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 671 GTAGTTATTGCTATTGGCTACCGTCAGCGATCTCTCCACCAAAGCTGCGTGCCGACCA 730

Qy 241 TGGGTGAAGCTCACACGAGAAAAGAGCCGACCCCTGCCTTGTGCAAGCAAGGCGTCG 300  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 731 TGGGAGAAGCTCACATGACAAACGTGCTGACCCAGCTTGTGCAAGACAAGGAGTGG 790

Qy 301 TAGACAGAGGATGGGGAATGGATGCGGACTGTTGAAAGGGGAGCATTGACACATGTG 360  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 791 TGGACAGGGCTGGGCAACGGCTGCGGACTATTGGCAAAGGAAGCATTGACACATGCG 850

Qy 361 CAAAGTTGCCTGTACAACCAAGGCAACTGGTGGATTATCCAGAAGGAAAACATCAAGT 420  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 851 CCAAATTGCGCTGCTACCAAGGAATAGGAAGAACATCTTGAAGAGAAATATCAAGT 910

Qy 421 ACGAGGTTGCCATATTGTGCATGGCCGACGACTGTCGAATCACATGGCAATTATTCAA 480  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 911 ACGAAGTGGCCATTGGCCATGGACCAACTACTGTGGAGTCGCACGGAAACTACCCCA 970

Qy 481 CACAGATAGGGCTACCAAGCAGGAAGGTCAGCATAACTCCATCGGCACCATCCTACA 540  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 971 CACAGGTTGGAGCCACTCAGGCAGGGAGATTAGCATCAGCATCAGCATCCTGCAGGCTTCATACA 1030

Qy 541 CGCTGAAGTTGGGTGAGTATGGTGAGGTACAGTTGACTGTGAGGCCACGGTCAGGAATAG 600  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1031 CACTAAAGCTTGGAGAATATGGAGAGGTGACAGTGGACTGTGAACCACGGTCAGGGATTG 1090

Qy 601 ACACTAGCGCTTACTACGTTATGTCAGTGGTGCAGTCAGTGGCTTGGTTACCGAGAAT 660  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1091 ACACCAATGCATACTACGTGATGACTGTTGAACAAAGACGTTCTGGTCCATCGTGAGT 1150

Qy 661 GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG 720  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1151 GGTCATGGATCTCACCTCCCTGGAGCAGTGCTGGAAGTACTGTGTGGAGGAACAGAG 1210

Qy 721 AAACACTGATGGAGTTGAAGAACCTCATGCCACCAACAAATCTGCGTAGCTCTAGGGT 780  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1211 AGACGTTAATGGAGTTGAGGAACCACACGCCACGAAGCAGTGTGATAGCATTGGGCT 1270

Qy	781	CGCAGGAAGGTGCCTGACCCAAGCTCTGGCTGGAGCAATTCTGTTGAGTCAGCA 840
Db	1271	CACAAGAGGGAGCTCTGCATCAAGCTTGGCTGGAGCCATTCTGTGGAATTTCAGCA 1330
Qy	841	ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAAGTTGCAGC 900
Db	1331	ACACTGTCAAGTTGACGTCGGTCATTGAAGTGTAGAGTGAAGATGGAAAAATTGCAGT 1390
Qy	901	TGAAGGGAACAAACATATGGGTATGCTCAAAGCATTCAAATTGCTAGGACTCCGCTG 960
Db	1391	TGAAGGGAACAAACCTATGGCGTCTGTCAGGCTTCAAGTTCTGGGACTCCGCGAG 1450
Qy	961	ACACTGGTCATGGAACGGTGGTGTGGAACTGCAGTATAACCGAAAAGACGGGCCTTGCA 1020
Db	1451	ACACAGGTACGGCACTGTGGTGTGAAATTGCAGTACACTGGCACGGATGGACCTTGCA 1510
Qy	1021	AAGTGCCATTCTCTGTGGCTTCCCTGAACGACCTTACACCCGTTGGAAGGCTGGTGA 1080
Db	1511	AAGTCCTATCTCGTCAGTGGCTCATGAAACGACCTAACGCCAGTGGCAGATTGGTCA 1570
Qy	1081	CTGTGAATCCATTGTGTCTGGCTACGGCCAACCTGAAGGTTTGATTGAACCTGAAC 1140
Db	1571	CTGTCAACCCTTTGTTCAGTGGCCACGGCCAACGCTAAGGTCCTGATTGAATTGGAAC 1630
Qy	1141	CCCCGTTAGTGAACCTTACATCGTGGTGGGGAGAGGAGAACAGCAGATAAACCAACT 1200
Db	1631	CACCCTTGGAGACTCATACATAGTGGTGGCAGAGGAGAACACAGATCAATCACCATT 1690
Qy	1201	GGCACAAATCTGGGAGCAGTATTGAAAGGCTTCACCACTACACTCAGAGGAGCTAAC 1260
Db	1691	GGCACAAAGTCTGGAAGCAGCATTGGCAAAGCCTTACAACCAACCCCTCAAAGGAGCGAGA 1750
Qy	1261	GACTTGCAGCTTGGAGACACTGCCTGGATTTGGATCAGTCGGAGGGTTTCACCT 1320
Db	1751	GACTAGCCGCTCTAGGAGACACAGCTGGGACTTGGATCAGTTGGAGGGTGTTCACCT 1810
Qy	1321	CGGTAGGGAAAGCCATACACCAAGTTGGAGGAGCCTTAGATCACTCTTGGAGGGA 1380
Db	1811	CAGTTGGGAAGGCTGTCCATCAAGTGTTCGGAGGAGCATTCCGCTCACTTTGGAGGCA 1870
Qy	1381	TGTCTGGATCACACAGGGCTCTGGAGCTCTGCTGTGGATGGAAATTACGCC 1440
Db	1871	TGTCTGGATAACGCAAGGATTGCTGGGGCTCCTGTGATGGGATGGCATTATGCTC 1930
Qy	1441	GTGACAGGTCAATTGCTATGACGTTCCCTGGGGTGGAGGAGTCTGCTCTTCCCTCGG 1500
Db	1931	GTGATAGGTCCATAGCTCTCACGTTCTCGCAGTTGGAGGAGTCTGCTCTTCCCTCCG 1990
Qy	1501	TCAACGTCCATGCT 1514
Db	1991	TGAACGTGCATGCT 2004

## RESULT 15

AY963774

## LOCUS

AY963774

2004 bp

RNA

linear

VRL 04-SEP-2005

DEFINITION West Nile virus polyprotein gene, partial cds.  
 ACCESSION AY963774  
 VERSION AY963774.1 GI:63098701  
 KEYWORDS .  
 SOURCE West Nile virus (WNV)  
 ORGANISM West Nile virus  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 Flavivirus; Japanese encephalitis virus group.  
 REFERENCE 1 (bases 1 to 2004)  
 AUTHORS Elizondo-Quiroga,D., Davis,C.T., Fernandez-Salas,I.,  
 Escobar-Lopez,R., Velazco Olmos,D., Soto Gastelum,L.C., Aviles  
 Acosta,M., Elizondo-Quiroga,A., Gonzalez-Rojas,J.I., Contreras  
 Cordero,J.F., Guzman,H., Travassos da Rosa,A., Blitvich,B.J.,  
 Barret,A.D., Beaty,B.J. and Tesh,R.B.  
 TITLE West Nile Virus Isolation in Human and Mosquitoes, Mexico  
 JOURNAL Emerging Infect. Dis. 11 (9), 1449-1452 (2005)  
 REFERENCE 2 (bases 1 to 2004)  
 AUTHORS Elizondo-Quiroga,D., Davis,C., Fernandez-Salas,I.,  
 Escobar-Lopez,R., Velazco-Olmos,D., Soto-Gastelum,L.,  
 Aviles-Acosta,M., Elizondo-Quiroga,A., Gonzalez-Rojas,J.,  
 Contreras-Cordero,J., Guzman,H., Travassos da Rosa,A., Blitvich,B.,  
 Beaty,B., Barret,A. and Tesh,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-MAR-2005) Pathology, University of Texas Medical  
 Branch, 301 University Ave., Galveston, TX 77555, USA  
 FEATURES Location/Qualifiers  
 source 1. .2004  
 /organism="West Nile virus"  
 /mol\_type="genomic RNA"  
 /serotype="lineage I"  
 /isolation\_source="serum"  
 /specific\_host="Homo sapiens"  
 /db\_xref="taxon:11082"  
 /country="Mexico: Sonora"  
 CDS <1. .>2004  
 /codon\_start=1  
 /product="polyprotein"  
 /protein\_id="AAV32589.1"  
 /db\_xref="GI:63098702"  
 /translation="VTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVRAMDVGYM  
 CDDTITYECPVLSAGNDPEDIDCWCTKSAYVRYGRCKTRHSRRSRSLTQTHGES  
 TLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRVVFVLLL  
 LVAPAYSFNCLGMSNRFLEGVSAGTWDLVLEGDSCVTIMSKDKPTIDVKMMNMEAA  
 NLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWNGCGLFG  
 KGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQAGATQAGRF  
 SITPAAPSYTTLKGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWFMIDLNL  
 SSAGSTVWRNRTELMEFEEPHATQSVIALGSQEGALHQALAGAI PVEFSSNTVKLTS  
 GHLKCRVKMEKLQLKGTTYGVCASKFKLGT PADTGHGTVVLELQYTGTDPCKVPIS  
 SVASLNDLTPVGRIVTNPVFSVATANAKVIELEPFFGDSYIVVGRGEQQINHHWHK  
 SGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAFRSLFGGM  
 SWITQGLLGALLWMGINARDRIALTFLAVGGVLLFLSVNVHA"  
 mat\_peptide 466. .966  
 /product="pre-membrane protein; prM/M"  
 mat\_peptide 967. .>2004  
 /product="envelope protein"  
 ORIGIN

Query Match 65.6%; Score 997.2; DB 10; Length 2004;  
Best Local Similarity 78.7%; Pred. No. 0;  
Matches 1191; Conservative 0; Mismatches 323; Indels 0; Gaps 0;

Qy 1 CGGAATTCACTGTTAGGAATGAGCAACAGGGACTTCCTGGAGGGAGTGTCTG 60  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 491 CAGCTTACAGCTCACTGCCTGGAAATGAGCAACAGAGACTTCTTGGAAAGGAGTGTCTG 550

Qy 61 GAGCTACATGGGTTGATCTGGTACTGGAAGGAGACAGTTGTGTGACCATAATGTCAAAG 120  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 551 GAGCAACATGGGTGGATTGGTCTCGAAGGCAGCTGCGTGAATCATGTCTAAAG 610

Qy 121 ACAAGCCAACCATTGATGTCAAATGATGAACATGGAAGCAGCTAACATCGCAGATGTGC 180  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 611 ACAAGCCTACCATCGATGTGAAGATGATGAATATGGAGGCGGCCAACCTGGCAGAGGTCC 670

Qy 181 GTAGCTACTGCTACTTAGCTTCGGTCAGTGATCTGTCAACAAAAGCCGCGTGTCCAACCA 240  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 671 GCAGTTATTGCTATTGGCTACCGTCAGCGATCTCTCCACCAAAGCTGCGTGCCTGGACCA 730

Qy 241 TGGGTGAAGCTCACACGAGAAAAGAGCCGACCCGCCTTGTGCAAGCAAGGCGTCG 300  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 731 TGGGAGAAGCTCACATGACAAACGTGCTGACCCAGCTTGTGCAAGACAAGGAGTGG 790

Qy 301 TAGACAGAGGATGGGAAATGGATGCGGACTGTTGAAAGGGAGCATTGACACATGTG 360  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 791 TGGACAGGGGCTGGGCAACGGCTGCGGACTATTGGAAAGGAAGCATTGACACATGCG 850

Qy 361 CAAAGTTGCCTGTACAACCAAGGCAACTGGTGGATTATCCAGAAGGAAAACATCAAGT 420  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 851 CCAAGTTGCCTGCTCTACCAAGGCAATAGGAAGAACATCTTGAAGAGAAATATCAAGT 910

Qy 421 ACGAGGTTGCCATATTGTGCATGGCCGACGACTGTCGAATCACATGGCAATTATTCAA 480  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 911 ACGAAGTGGCCATTGGCATGGACCAACTACTGTGGAGTCGCACGGAAACTACTCCA 970

Qy 481 CACAGATAGGGCTACCAAGCAGGAAGGTCAGCATAACTCCATCGGCACCATCCTACA 540  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 971 CACAGGCTGGAGCCACTCAGGCAGGGAGATTAGCATCAGTGCCTGGCGCCTCATACA 1030

Qy 541 CGCTGAAGTTGGGTGAGTATGGTGGAGTCACAGTTGACTGTGAGGCCACGGTCAGGAATAG 600  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1031 CACTAAAGCTGGAGAATATGGAGAGGTGACAGTGGACTGTGAACCACGGTCAGGGATTG 1090

Qy 601 ACACTAGCGCTTACTACGTTATGTCAGTGGTGCAGTCAGTCCTTGGTTACCGAGAAT 660  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1091 ACACCAATGCATACTACGTGATGACTGTTGAAACAAAGACGTTCTGGTCCATCGTGAGT 1150

Qy 661 GGTTTATGGACCTGAACCTTCCATGGAGTAGCGCTGGAAGCACAACGTGGAGGAACCGGG 720  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1151 GGTTCATGGACCTCAACCTCCCTGGAGCAGTGTGAGCAGTGTGTGGAGGAACAGAG 1210

Qy 721 AACACTGATGGAGTTGAAGAACCTCATGCCACCAAAACATCTGCGTAGCTCTAGGGT 780  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1211 AGACGTTAATGGAGTTGAGGAACCAACGCCACGAAGCAGTGTGATAGCATTGGGCT 1270

Qy 781 CGCAGGAAGGTGCCTGCACCAAGCTGGCTGGAGCAATTCTGTTGAGTTCTCAAGCA 840

Db	1271 CACAAGAGGGAGCTCTGCATCAAGCTTGGCTGGAGCCATTCTGTGGAATTTCAGCA 1330
Qy	841 ACACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTAAGATGGAGAAGTTGCAGC 900 
Db	1331 ACACTGTCAAGTTGACGTCGGGTCAATTGAAGTGTAGAGTGAAGATGGAAAAATTGCAGT 1390
Qy	901 TGAAGGGAAACAACATATGGTGTATGCTCAAAGCATTCAAATTCGCTAGGACTCCGCTG 960 
Db	1391 TGAAGGGAAACAACCTATGGCGTCTGTTCAAAGGCTTCAAGTTCTGGACTCCGAG 1450
Qy	961 ACACTGGTCATGGAACGGTGGTGGAACTGCAGTATACCGAAAAGACGGGCCTTGCA 1020 
Db	1451 ACACAGGTACGGCACTGTGGTGGAAATTGCAGTACACTGGCACGGATGGACCTTGTG 1510
Qy	1021 AAGTGCCCATTCTCTGTGGCTCCCTGAACGACCTAACCCGGTGGAAAGGCTGGTGA 1080 
Db	1511 AAGTTCCATCTCGTCAGTGGCTCATTGAACGACCTAACGCCAGTGGCAGATTGGTCA 1570
Qy	1081 CTGTGAATCCATTGTGTCTGTGGCTACGGCCAECTCGAAGGTTTGATTGAACCTCGAAC 1140 
Db	1571 CTGTCAACCCTTTGTTCAAGTGGCACGCCAACGCTAACGGTCTGATTGAATTGGAAC 1630
Qy	1141 CCCCGTTTAGTGAECTTACATCGTGGTGGGGAGAGGAGAACAGCAGATAAACCAACT 1200   .
Db	1631 CACCCCTTGAGACTCATACATAGTGGTGGGCAGAGGAGAACACAGATCAATCACCATT 1690
Qy	1201 GGCACAAATCTGGGAGCAGTATTGGAAAGGCTTCACCACTACACTCAGAGGAGCTAAC 1260 
Db	1691 GGCACAAAGTCTGGAAGCAGCATTGGCAAAGCCTTACAACCACCTCAAAGGAGCGCAGA 1750
Qy	1261 GACTTGCAGCTCTGGAGACACTGCCTGGATTTGGATCAGTCGGAGGGTTTCACCT 1320 
Db	1751 GACTAGCCGCTCTAGGAGACACAGCTGGACTTTGGATCAGTTGGAGGGGTGTTCACCT 1810
Qy	1321 CGGTAGGGAAAGCCATACACCAAGTTTGGAGGAGCCTTAGATCACTCTTGGAGGG 1380 
Db	1811 CAGTTGGGAAGGCTGTCCATCAAGTGTTCGGAGGAGCATTCCGCTCACTGTTGGAGGGCA 1870
Qy	1381 TGTCCTGGATCACACAGGGCTTCTGGAGCTCTCTGCTGTGGATGGAAATTACGCC 1440 
Db	1871 TGTCCTGGATAACGCAAGGATTGCTGGGGCTCTCCTGTTGGATGGCATCAATGCTC 1930
Qy	1441 GTGACAGGTCAATTGCTATGACGTTCTGGCGTTGGAGGAGTCTGCTCTCCCTTCGG 1500 
Db	1931 GTGATAGGTCCATAGCTCTCACGTTCTCGCAGTTGGAGGAGTCTGCTCTCCCTCCG 1990
Qy	1501 TCAACGTCCATGCT 1514 
Db	1991 TGAACGTGCATGCT 2004

Search completed: June 12, 2006, 19:10:35  
 Job time : 8801 secs

DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
 DR GO; GO:0003725; F:double-stranded RNA binding; IEA.  
 DR GO; GO:0003724; F:RNA helicase activity; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR011545; DEAD/DEAH\_N.  
 DR InterPro; IPR011999; Flav\_glyE\_cen\_dm.  
 DR InterPro; IPR001122; Flavi\_capsidC.  
 DR InterPro; IPR011492; Flavi\_DEAD.  
 DR InterPro; IPR000069; Flavi\_M.  
 DR InterPro; IPR001157; Flavi\_NS1.  
 DR InterPro; IPR000752; Flavi\_NS2A.  
 DR InterPro; IPR000487; Flavi\_NS2B.  
 DR InterPro; IPR000404; Flavi\_NS4A.  
 DR InterPro; IPR001528; Flavi\_NS4B.  
 DR InterPro; IPR000208; Flavi\_NS5.  
 DR InterPro; IPR002535; Flavi\_propep.  
 DR InterPro; IPR000336; Flv\_glyE\_Ig-like.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR001850; Peptidase\_S7.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR InterPro; IPR002877; RrmJFtsJ\_mtfrase.  
 DR InterPro; IPR011998; Vrl\_glyE\_cen\_dim.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF01003; Flavi\_capsid; 1.  
 DR Pfam; PF07652; Flavi\_DEAD; 1.  
 DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
 DR Pfam; PF00869; Flavi\_glycoprot; 1.  
 DR Pfam; PF01004; Flavi\_M; 1.  
 DR Pfam; PF00948; Flavi\_NS1; 1.  
 DR Pfam; PF01005; Flavi\_NS2A; 1.  
 DR Pfam; PF01002; Flavi\_NS2B; 1.  
 DR Pfam; PF01350; Flavi\_NS4A; 1.  
 DR Pfam; PF01349; Flavi\_NS4B; 1.  
 DR Pfam; PF00972; Flavi\_NS5; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 DR Pfam; PF01728; FtsJ; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00949; Peptidase\_S7; 1.  
 DR ProDom; PD001496; Flavi\_NS1; 1.  
 DR SMART; SM00487; DEXDc; 1.  
 DR SMART; SM00490; HELICc; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 KW Polyprotein.  
 FT CHAIN 1 123 capsid.  
 FT CHAIN 124 215 prM.  
 FT CHAIN 216 290 M.  
 FT CHAIN 291 791 envelope.  
 FT CHAIN 792 1144 NS1.  
 FT CHAIN 1145 1374 NS2A.  
 FT CHAIN 1375 1505 NS2B.

FT CHAIN 1506 2124 NS3.  
 FT CHAIN 2125 2273 NS4A.  
 FT CHAIN 2274 2528 NS4B.  
 FT CHAIN 2529 3433 NS5.  
 SQ SEQUENCE 3433 AA; 381124 MW; C302F24541A66BC8 CRC64;  
  
 Query Match 95.9%; Score 2531; DB 2; Length 3433;  
 Best Local Similarity 95.4%; Pred. No. 3.8e-183;  
 Matches 478; Conservative 14; Mismatches 9; Indels 0; Gaps 0;  
  
 Qy 1 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 60  
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 Db 291 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYC 350  
  
 Qy 61 YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWGNCGGLFGKGSIDTCAKFA 120  
 |||:|||||:|||||:|||||:|||||:  
 Db 351 YLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWGNCGGLFGKGSIDTCAKFA 410  
  
 Qy 121 CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL 180  
 |:|||||:|||:|||||:|||||:|||||:|||||:  
 Db 411 CSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPAAPSYTLKL 470  
  
 Qy 181 GEYGEVTVDCEPRSGIDTSAYYVMSVGAFLVHREWFDLNLPWSSAGSTWRNRETL 240  
 |||||||:|||||:|||:|||:|||||:  
 Db 471 GEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWFDLNLPWSSAGSTVWRNRETL 530  
  
 Qy 241 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVMEKLQLKGT 300  
 |||||||:|||||:|||||:|||||:  
 Db 531 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVMEKLQLKGT 590  
  
 Qy 301 TYGVCSKAFKFARTPDTGHGTVVLELQYTGKDGPCKVPISSVASLNDLTPVGRLVNP 360  
 |||||||:|||||:  
 Db 591 TYGVCSKAFKFLGTPDTGHGTVVLELQYTGDGPCKVPISSVASLNDLTPVGRLVNP 650  
  
 Qy 361 FVSVATANSKVLIELEPPFSDSYIVVGRGEQQINHHWHKGSSIGKAFTTLRGAQRLAA 420  
 |||||||:|||||:  
 Db 651 FVSVATANAKVLIELEPPFGDSYIVVGRGEQQINHHWHKGSSIGKAFTTLKGAQRLAA 710  
  
 Qy 421 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARD 480  
 |||||||:|||||:  
 Db 711 LGDTAWDFGSVGGVFTSVGKAVHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARD 770  
  
 Qy 481 IAMTFLAVGGVLLFLSVNVHA 501  
 ||:|||||:  
 Db 771 IALTFLAVGGVLLFLSVNVHA 791

RESULT 13  
 Q8JU42\_WNV  
 ID Q8JU42\_WNV PRELIMINARY; PRT; 3433 AA.  
 AC Q8JU42;  
 DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.  
 DT 01-OCT-2002, sequence version 1.  
 DT 07-FEB-2006, entry version 14.  
 DE Polyprotein.  
 OS West Nile virus (WN).  
 OC Viruses; ssRNA positive-strand viruses; no DNA stage; Flaviviridae;

OC Flavivirus; Japanese encephalitis virus group.  
OX NCBI\_TaxID=11082;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22089180; PubMed=12093177; DOI=10.1006/viro.2002.1449;  
RA Lanciotti R.S., Ebel G.D., Deubel V., Kerst A.J., Murri S., Meyer R.,  
RA Bowen M., McKinney N., Morrill W.E., Crabtree M.B., Kramer L.D.,  
RA Roehrig J.T.;  
RT "Complete genome sequences and phylogenetic analysis of West Nile  
RT virus strains isolated from the United States, Europe, and the Middle  
RT East.";  
RL Virology 298:96-105(2002).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AF404757; AAM81753.1; -; Genomic\_RNA.  
DR HSSP; Q88653; 1L9K.  
DR SMR; Q8JU42; 25-97.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003725; F:double-stranded RNA binding; IEA.  
DR GO; GO:0003724; F:RNA helicase activity; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH\_N.  
DR InterPro; IPR011999; Flav\_glyE\_cen\_dm.  
DR InterPro; IPR001122; Flavi\_capsidC.  
DR InterPro; IPR011492; Flavi\_DEAD.  
DR InterPro; IPR000069; Flavi\_M.  
DR InterPro; IPR001157; Flavi\_NS1.  
DR InterPro; IPR000752; Flavi\_NS2A.  
DR InterPro; IPR000487; Flavi\_NS2B.  
DR InterPro; IPR000404; Flavi\_NS4A.  
DR InterPro; IPR001528; Flavi\_NS4B.  
DR InterPro; IPR000208; Flavi\_NS5.  
DR InterPro; IPR002535; Flavi\_propep.  
DR InterPro; IPR000336; Flv\_glyE\_Ig-like.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR001850; Peptidase\_S7.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR InterPro; IPR002877; RrmJFtsJ\_mtfrase.  
DR InterPro; IPR011998; Vrl\_glyE\_cen\_dim.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF01003; Flavi\_capsid; 1.  
DR Pfam; PF07652; Flavi\_DEAD; 1.  
DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
DR Pfam; PF00869; Flavi\_glycoprot; 1.  
DR Pfam; PF01004; Flavi\_M; 1.  
DR Pfam; PF00948; Flavi\_NS1; 1.

DR Pfam; PF01005; Flavi\_NS2A; 1.  
 DR Pfam; PF01002; Flavi\_NS2B; 1.  
 DR Pfam; PF01350; Flavi\_NS4A; 1.  
 DR Pfam; PF01349; Flavi\_NS4B; 1.  
 DR Pfam; PF00972; Flavi\_NS5; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 DR Pfam; PF01728; FtsJ; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00949; Peptidase\_S7; 1.  
 DR ProDom; PD001496; Flavi\_NS1; 1.  
 DR SMART; SM00487; DEXDc; 1.  
 DR SMART; SM00490; HELICc; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 KW Polyprotein.  
 SQ SEQUENCE 3433 AA; 381210 MW; 1DFFCCDB2174B7EE CRC64;  
  
 Query Match 95.9%; Score 2531; DB 2; Length 3433;  
 Best Local Similarity 95.4%; Pred. No. 3.8e-183;  
 Matches 478; Conservative 14; Mismatches 9; Indels 0; Gaps 0;  
  
 Qy 1 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 60  
 |||||||  
 Db 291 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYC 350  
  
 Qy 61 YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWGNCGLFGKGSIDTCAKFA 120  
 |||:  
 Db 351 YLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWGNCGLFGKGSIDTCAKFA 410  
  
 Qy 121 CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL 180  
 |:  
 Db 411 CSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPAAPSYTLKL 470  
  
 Qy 181 GEYGEVTVDCEPRSGIDTSAYVMSVGAKSFLVHREWMDLNLPWSSAGSTTWRNRETL 240  
 |||||||:  
 Db 471 GEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWMDLNLPWSSAGSTVWRNRETL 530  
  
 Qy 241 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKEKLQLKGT 300  
 |||||||:  
 Db 531 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKEKLQLKGT 590  
  
 Qy 301 TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPCKVPISSVASLNDLTPVGRVLTVNP 360  
 |||||||  
 Db 591 TYGVCSKAFKFLGT PADTGHGTVVLELQYTGTDPCKVPISSVASLNDLTPVGRVLTVNP 650  
  
 Qy 361 FVSVATANSKVLIIELEPPFSDSYIVVGRGEQQINHHWHKGSSSIGKAFTTLRGAQRLLAA 420  
 |||||||:  
 Db 651 FVSVATANAKVLIIELEPPFGDSYIVVGRGEQQINHHWHKGSSSIGKAFTTLKGAAQRLLAA 710  
  
 Qy 421 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLSFGGMSWITQGLLGALLWMGINARDRS 480  
 |||||||:  
 Db 711 LGDTAWDFGSVGGVFTSVGKAVHQVFGGAFRSLSFGGMSWITQGLLGALLWMGINARDRS 770  
  
 Qy 481 IAMTFLAVGGVLLFLSVNVHA 501  
 ||:  
 Db 771 IALTFLAVGGVLLFLSVNVHA 791

RESULT 14

Q9EA21\_WNV

ID Q9EA21\_WNV PRELIMINARY; PRT; 3433 AA.

AC Q9EA21;

DT 01-MAR-2001, integrated into UniProtKB/TrEMBL.

DT 01-MAR-2001, sequence version 1.

DT 07-FEB-2006, entry version 20.

DE Polyprotein.

OS West Nile virus (WN).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Flavivirus; Japanese encephalitis virus group.

OX NCBI\_TaxID=11082;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=RO97-50;

RX MEDLINE=20014331; PubMed=10548295;

RA Savage H.M., Ceianu C., Nicolescu G., Karabatsos N., Lanciotti R.,  
RA Vladimirescu A., Laiv L., Ungureanu A., Romanca C., Tsai T.F.;

RT "Entomologic and avian investigations of an epidemic of West Nile  
RT fever in Romania in 1996, with serologic and molecular  
RT characterization of a virus isolate from mosquitoes.";

RL Am. J. Trop. Med. Hyg. 61:600-611(1999).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=RO97-50;

RA Bowen M., Meyer R.F., McKinney N., Morrill W., Lanciotti R.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.

CC -----

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL; AF260969; AAG02040.1; -; Genomic\_RNA.

DR HSSP; Q88653; 1L9K.

DR SMR; Q9EA21; 25-97.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.

DR GO; GO:0003725; F:double-stranded RNA binding; IEA.

DR GO; GO:0003724; F:RNA helicase activity; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0019079; P:viral genome replication; IEA.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR011545; DEAD/DEAH\_N.

DR InterPro; IPR011999; Flav\_glyE\_cen\_dm.

DR InterPro; IPR001122; Flavi\_capsidC.

DR InterPro; IPR011492; Flavi\_DEAD.

DR InterPro; IPR000069; Flavi\_M.

DR InterPro; IPR001157; Flavi\_NS1.

DR InterPro; IPR000752; Flavi\_NS2A.

DR InterPro; IPR000487; Flavi\_NS2B.

DR InterPro; IPR000404; Flavi\_NS4A.

DR InterPro; IPR001528; Flavi\_NS4B.

DR InterPro; IPR000208; Flavi\_NS5.

DR InterPro; IPR002535; Flavi\_propep.  
 DR InterPro; IPR000336; Flv\_glyE\_Ig-like.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR001850; Peptidase\_S7.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR InterPro; IPR002877; RrmJFtsJ\_mtfrase.  
 DR InterPro; IPR011998; Vrl\_glyE\_cen\_dim.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF01003; Flavi\_capsid; 1.  
 DR Pfam; PF07652; Flavi\_DEAD; 1.  
 DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
 DR Pfam; PF00869; Flavi\_glycoprot; 1.  
 DR Pfam; PF01004; Flavi\_M; 1.  
 DR Pfam; PF00948; Flavi\_NS1; 1.  
 DR Pfam; PF01005; Flavi\_NS2A; 1.  
 DR Pfam; PF01002; Flavi\_NS2B; 1.  
 DR Pfam; PF01350; Flavi\_NS4A; 1.  
 DR Pfam; PF01349; Flavi\_NS4B; 1.  
 DR Pfam; PF00972; Flavi\_NS5; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 DR Pfam; PF01728; FtsJ; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00949; Peptidase\_S7; 1.  
 DR ProDom; PD001496; Flavi\_NS1; 1.  
 DR SMART; SM00487; DEXDc; 1.  
 DR SMART; SM00490; HELICc; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.

KW Polyprotein.

FT	CHAIN	1	123	nucleocapsid protein C.
FT	CHAIN	124	215	pre-membrane protein prM.
FT	CHAIN	216	290	membrane protein M.
FT	CHAIN	291	791	envelope glycoprotein E.
FT	CHAIN	792	1143	non-structural protein 1 NS1.
FT	CHAIN	1144	1374	non-structural protein 2A NS2A.
FT	CHAIN	1375	1505	non-structural protein 2B NS2B.
FT	CHAIN	1506	2124	non-structural protein 3 NS3.
FT	CHAIN	2125	2273	non-structural protein 4A NS4A.
FT	CHAIN	2274	2528	non-structural protein NS4B.
FT	CHAIN	2529	3433	non-structural protein NS5.
SQ	SEQUENCE	3433 AA;	381256 MW;	4695F8911670DF2A CRC64;

Query Match 95.9%; Score 2531; DB 2; Length 3433;  
 Best Local Similarity 95.4%; Pred. No. 3.8e-183;  
 Matches 478; Conservative 14; Mismatches 9; Indels 0; Gaps 0;

Qy	1	FNCLGMSNRDFLEGVSGATWVDLVLEGDCVTIMSKDKPTIDVKMMNMEAANLADVR SYC	60
		:     :     :     :     :     :     :	
Db	291	FNCLGMSNRDFLEGVSGATWVDLVLEGDCVTIMSKDKPTIDVKMMNMEAANLAEVR SYC	350
Qy	61	YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWGNCGGLFGKGSIDTCAKFA	120
		:     :     :     :     :     :     :     :     :     :	
Db	351	YLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWGNCGGLFGKGSIDTCAKFA	410
Qy	121	CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPS YTLKL	180
		:     :     :     :     :     :     :     :     :	
Db	411	CSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPAAPS YTLKL	470

Qy 181 GEYGEVTVDCEPRSGIDTSAYYMSVGAKSFLVHREWFMDLNLPWSSAGSTTWRNRETL 240  
 |||||||:|||||:||| :|||||:||| :|||||:|||||:|||||:|||||  
 Db 471 GEYGEVTVDCEPRSGIDTNAYYMTVGTKTFLVHREWFMDLNLPWSSAGSTVWRNRETL 530  
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 Qy 241 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 300  
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 Db 531 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 590  
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 Qy 301 TYGVCSKAFKFARTPADTGHTVVLELQYTGKDGPCKVPISSVASLNDLTPVGRLLTVNP 360  
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 Db 591 TYGVCSKAFKFLGTPADTGHTVVLELQYTGKDGPCKVPISSVASLNDLTPVGRLLTVNP 650  
 |||||||:|||||:|||||:|||||:|||||:|||||:  
 Qy 361 FVSVATANSKVLIIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLRGAQRLAA 420  
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 Db 651 FVSVATANAKVLIIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLKGAQRLAA 710  
 |||||||:|||||:  
 Qy 421 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS 480  
 |||||||:|||||:|||||:|||||:|||||:  
 Db 711 LGDTAWDFGSVGGVFTSVGKAVHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS 770  
 |||||||:  
 Qy 481 IALTFLAVGGVLLFLSVNVHA 501  
 ||:|||||:  
 Db 771 IALTFLAVGGVLLFLSVNVHA 791

#### RESULT 15

**Q9WHD2\_WNV**  
 ID Q9WHD2\_WNV PRELIMINARY; PRT; 773 AA.  
 AC Q9WHD2;  
 DT 01-NOV-1999, integrated into UniProtKB/TrEMBL.  
 DT 01-NOV-1999, sequence version 1.  
 DT 07-FEB-2006, entry version 24.  
 DE Polyprotein (Fragment).  
 OS West Nile virus (WN).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus; Japanese encephalitis virus group.  
 OX NCBI\_TaxID=11082;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=96-1030;  
 RX MEDLINE=98407299; PubMed=9737281; DOI=10.1016/S0140-6736(98)03538-7;  
 RA Tsai T.F., Popovici F., Cernescu C., Campbell G.L., Nedelcu N.I.;  
 RT "West Nile encephalitis epidemic in southeastern Romania.";  
 RL Lancet 352:767-771(1998).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=96-1030;  
 RA Lanciotti R.L., Ludwig M.L., Savage H.M.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; AF130363; AAD28624.1; -; Genomic\_RNA.  
 DR HSSP; Q88653; 10KE.  
 DR SMR; Q9WHD2; 1-72.

DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0019058; P:viral infectious cycle; IEA.  
 DR InterPro; IPR011999; Flav\_glyE\_cen\_dm.  
 DR InterPro; IPR001122; Flavi\_capsidC.  
 DR InterPro; IPR000069; Flavi\_M.  
 DR InterPro; IPR002535; Flavi\_propep.  
 DR InterPro; IPR000336; Flv\_glyE\_Ig-like.  
 DR InterPro; IPR011998; Vrl\_glyE\_cen\_dim.  
 DR Pfam; PF01003; Flavi\_capsid; 1.  
 DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
 DR Pfam; PF00869; Flavi\_glycoprot; 1.  
 DR Pfam; PF01004; Flavi\_M; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 KW Polyprotein.  
 FT CHAIN <1 88 capsid protein.  
 FT CHAIN 89 265 pre-membrane/membrane protein.  
 FT CHAIN 266 766 envelope glycoprotein.  
 FT NON\_TER 1 1  
 FT NON\_TER 773 773  
 SQ SEQUENCE 773 AA; 83362 MW; 2960B1E9AF064BF6 CRC64;  
  
 Query Match 95.9%; Score 2529; DB 2; Length 773;  
 Best Local Similarity 95.4%; Pred. No. 7.3e-184;  
 Matches 478; Conservative 13; Mismatches 10; Indels 0; Gaps 0;  
  
 Qy 1 FNCLGMSNRDFLEGVSGATWVDLVLEGSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 60  
 |||||||  
 Db 266 FNCLGMSNRDFLEGVSGATWVDLVLEGSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYC 325  
  
 Qy 61 YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWGNCGLFGKGSIDTCAKFA 120  
 ||:|||||  
 Db 326 YLATVSDLSTKAACPTMGEAHNDKRADPAFVCKQGVVDRGWGNCGLFGKGSIDTCAKFA 385  
  
 Qy 121 CTTKATGWIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL 180  
 |:|||||  
 Db 386 CSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYPTQIGATQAGRFSITPAAPSYTLKL 445  
  
 Qy 181 GEYGEVTVDCEPRSGIDTSAYYMSVGAKSFLVHREWMDLNLPWSSAGSTWRNRETLIM 240  
 |||||||:  
 Db 446 GEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWMDLNLPWSSAGSTVWRNRETLIM 505  
  
 Qy 241 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 300  
 |||||||:  
 Db 506 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 565  
  
 Qy 301 TYGVCSKAFKFARTPADTGHTVVLELQYTGKDGPCKVPISSVASLNDTPVGRLVTVNP 360  
 |||||||  
 Db 566 TYGVCSKAFKFLGTPADTGHTVVLELQYTGDGPCKVPISSVASLNDTPVGRLVTVNP 625  
  
 Qy 361 FVSVATANSKVVLIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLRGAQRILAA 420  
 |||||||:  
 Db 626 FVSVATANAKVVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLKGAQRILAA 685  
  
 Qy 421 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS 480

Db           |||||||||||||||||:||||||||||||||||||||||||||  
686 LGDTAWDFGSVGGVFTSVGKAVHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS 745  
  
Qy           481 IAMTFLAVGGVLLFLSVNVHA 501  
          ||:|||||||||||||||  
Db           746 IALTFLAVGGVLLFLSVNVHA 766

Search completed: June 10, 2006, 02:43:07  
Job time : 303 secs

RC STRAIN=ArB3573/82;  
 RA Borisevich V.G., Seregin A.V., Yamshchikov V.F.;  
 RT "Genetic determinants of West Nile virus pathogenicity.";  
 RL Submitted (DEC-2005) to the EMBL/GenBank/DDBJ databases.  
 CC  
 Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 DR EMBL; DQ318020; ABC49717.1; -; mRNA.  
 KW Polyprotein; Signal.  
 FT SIGNAL 106 123 Potential.  
 FT SIGNAL 290 792 Potential.  
 FT SIGNAL 2251 2273 Potential.  
 FT CHAIN 1 105 C protein.  
 FT CHAIN 106 290 prM protein.  
 FT CHAIN 124 215 cleaved amino terminal prM fragment.  
 FT CHAIN 216 290 M protein.  
 FT CHAIN 291 791 E protein.  
 FT CHAIN 792 1143 NS1 protein.  
 FT CHAIN 1144 1374 NS2A protein.  
 FT CHAIN 1375 1505 NS2B protein.  
 FT CHAIN 1506 2124 NS3 protein.  
 FT CHAIN 2125 2273 NS4A protein.  
 FT CHAIN 2274 2529 NS4B protein.  
 FT CHAIN 2530 3434 NS5 protein.  
 SQ SEQUENCE 3434 AA; 380337 MW; DF4C043FCA4F25DE CRC64;  
  
 Query Match 98.5%; Score 2599; DB 2; Length 3434;  
 Best Local Similarity 98.8%; Pred. No. 2.5e-188;  
 Matches 495; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
 Qy 1 FNCLGMSNRDFLEGVSGATWVLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVR SYC 60  
 |||||||  
 Db 291 FNCLGMSNRDFLEGVSGATWVLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVR SYC 350  
  
 Qy 61 YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA 120  
 |||||:  
 Db 351 YLASVSDLSTRAACPTMGEAHNEKRADPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA 410  
  
 Qy 121 CTTKATGWIIQKENIKYEVAI FVHGPTTVESHGNYSTQIGATQAGRFSITPSAPS YTLKL 180  
 |||||:  
 Db 411 CTTKATGWIIQKENIKYEVAI FVHGPTTVESHGDYSTQIGATQAGRFSITPSAPS YTLKL 470  
  
 Qy 181 GEYGEVTVDCEPRSGIDTSAYYVMSVGAKSFLVHREWMDLNL PWSSAGSTTWRNRE TL 240  
 |||||:  
 Db 471 GEYGEVTVDCEPRSGIDTSAYYVMSVGAKSFLVHREWMDLNL PWSSAGSTTWRNRE TL 530  
  
 Qy 241 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 300  
 |||||:  
 Db 531 EFEEPHATKRSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 590  
  
 Qy 301 TYGVCSKAFKFARTP ADTGHGTVVLELQYTGKDGPCKVPISSVASLNDLTPVGR LTVNP 360  
 |||||:  
 Db 591 TYGVCSKAFKFAGTPADTGHGTVVLELQYTGDGPCKVPISSVASLNDLTPVGR LTVNP 650  
  
 Qy 361 FVSVATANSKV LIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTT LRG AQR LAA 420  
 |||||:

Db 651 FVSVATANSKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLRGAQRLAA 710  
Qy 421 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS 480  
|||  
Db 711 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS 770  
|||  
Qy 481 IAMTFLAVGGVLLFLSVNVHA 501  
|||  
Db 771 IAMTFLAVGGVLLFLSVNVHA 791

RESULT 4

Q5MXE3\_WNV

ID Q5MXE3\_WNV PRELIMINARY; PRT; 3430 AA.  
AC Q5MXE3;  
DT 01-FEB-2005, integrated into UniProtKB/TrEMBL.  
DT 01-FEB-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Polyprotein.  
OS West Nile virus (WN).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Flavivirus; Japanese encephalitis virus group.  
OX NCBI\_TaxID=11082;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=B956;  
RX PubMed=15527855; DOI=10.1016/j.virol.2004.09.014;  
RA Yamshchikov G., Borisevich V., Seregin A., Chaporgina E., Mishina M.,  
RA Mishin V., Wai Kwok C., Yamshchikov V.;  
RT "An attenuated West Nile prototype virus is highly immunogenic and  
RT protects against the deadly NY99 strain: a candidate for live WN  
RT vaccine development.";  
RL Virology 330:304-312(2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=B956;  
RA Borisevich V.G., Yamshchikov V.F.;  
RT "Molecular basis of attenuation of the West Nile virus prototype  
RT strain B956.";  
RL Submitted (JAN-2004) to the EMBL/GenBank/DDBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AY532665; AAT02759.1; -; Genomic\_RNA.  
DR SMR; Q5MXE3; 25-97.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003725; F:double-stranded RNA binding; IEA.  
DR GO; GO:0003724; F:RNA helicase activity; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.

DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR011545; DEAD/DEAH\_N.  
 DR InterPro; IPR011999; Flav\_glyE\_cen\_dm.  
 DR InterPro; IPR001122; Flavi\_capsidC.  
 DR InterPro; IPR011492; Flavi\_DEAD.  
 DR InterPro; IPR000069; Flavi\_M.  
 DR InterPro; IPR001157; Flavi\_NS1.  
 DR InterPro; IPR000752; Flavi\_NS2A.  
 DR InterPro; IPR000487; Flavi\_NS2B.  
 DR InterPro; IPR000404; Flavi\_NS4A.  
 DR InterPro; IPR001528; Flavi\_NS4B.  
 DR InterPro; IPR000208; Flavi\_NS5.  
 DR InterPro; IPR002535; Flavi\_propep.  
 DR InterPro; IPR000336; Flv\_glyE\_Ig-like.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR001850; Peptidase\_S7.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR InterPro; IPR002877; RrmJFtsJ\_mtfrase.  
 DR InterPro; IPR011998; Vrl\_glyE\_cen\_dim.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF01003; Flavi\_capsid; 1.  
 DR Pfam; PF07652; Flavi\_DEAD; 1.  
 DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
 DR Pfam; PF00869; Flavi\_glycoprot; 1.  
 DR Pfam; PF01004; Flavi\_M; 1.  
 DR Pfam; PF00948; Flavi\_NS1; 1.  
 DR Pfam; PF01005; Flavi\_NS2A; 1.  
 DR Pfam; PF01002; Flavi\_NS2B; 1.  
 DR Pfam; PF01350; Flavi\_NS4A; 1.  
 DR Pfam; PF01349; Flavi\_NS4B; 1.  
 DR Pfam; PF00972; Flavi\_NS5; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 DR Pfam; PF01728; FtsJ; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00949; Peptidase\_S7; 1.  
 DR ProDom; PD001496; Flavi\_NS1; 1.  
 DR SMART; SM00487; DEXDc; 1.  
 DR SMART; SM00490; HELICc; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 KW Polyprotein.  
 SQ SEQUENCE 3430 AA; 379894 MW; 6298C302480200D8 CRC64;  
  
 Query Match 97.6%; Score 2575; DB 2; Length 3430;  
 Best Local Similarity 98.2%; Pred. No. 1.7e-186;  
 Matches 492; Conservative 3; Mismatches 2; Indels 4; Gaps 1;  
  
 Qy 1 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVR SYC 60  
 |||||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:  
 Db 291 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTLMSKDKPTIDVKMMNMEAANLADVR SYC 350  
 |||||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:  
 Qy 61 YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA 120  
 |||||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:  
 Db 351 YLASVSDLSTRAACPTMGEAHNEKRADPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA 410  
 |||||||:||||||:||||||:||||||:||||||:||||||:  
 Qy 121 CTTKATGWIIQKENIKYEVAFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYT LKL 180  
 |||||||:||||||:||||||:||||||:||||||:

Db	411	CTTKATGWIIQKENIKYEVAIFVHGPTTVESHG---	KIGATQAGRFSITPSAPSYTTLKL	466
Qy	181	GEYGEVTVDCEPRSGIDTSAYYMSVGAKSFLVHREWFMDLNLPWSSAGSTWRNRETLM		240
Db	467	GEYGEVTVDCEPRSGIDTSAYYMSVGAKSFLVHREWFMDLNLPWSSAGSTWRNRETLM		526
Qy	241	EFEPEPHATQSVVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT		300
Db	527	EFEPEPHATQSVVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT		586
Qy	301	TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPCKVPISSVASLNDLTPVGRLVTVNP		360
Db	587	TYGVCSKAFKFARTPADTGHGTVVLELQYTGDGPCKVPISSVASLNDLTPVGRLVTVNP		646
Qy	361	FVSVATANSKVLELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLRGAQRLAA		420
Db	647	FVSVATANSKVLELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLRGAQRLAA		706
Qy	421	LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLSFGGMSWITQGLLGALLWMGINARDRS		480
Db	707	LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLSFGGMSWITQGLLGALLWMGINARDRS		766
Qy	481	IAMTFLAVGGVLLFLSVNVHA	501	
Db	767	IAMTFLAVGGVLLFLSVNVHA	787	

## RESULT 5

Q2PMF5\_WNV

ID Q2PMF5\_WNV PRELIMINARY; PRT; 3430 AA.

AC Q2PMF5;

DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.

DT 24-JAN-2006, sequence version 1.

DT 07-FEB-2006, entry version 2.

DE Polyprotein precursor.

OS West Nile virus (WN).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Flavivirus; Japanese encephalitis virus group.

OX NCBI\_TaxID=11082;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=ArD76104;

RA Borisevich V.G., Seregin A.V., Yamshchikov V.F.;

RT "Genetic determinants of West Nile virus pathogenicity.";

RL Submitted (DEC-2005) to the EMBL/GenBank/DDBJ databases.

CC -----

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL; DQ318019; ABC49716.1; -; mRNA.

KW Polyprotein; Signal.

FT SIGNAL 106 123 Potential.

FT SIGNAL 275 290 Potential.

FT SIGNAL 764 787 Potential.

FT SIGNAL 2247 2269 Potential.

FT CHAIN 1 105 C protein.

FT CHAIN 124 215 cleaved amino terminal prM fragment.

FT CHAIN 124 290 prM protein.  
 FT CHAIN 216 290 M protein.  
 FT CHAIN 291 787 E protein.  
 FT CHAIN 788 1139 NS1 protein.  
 FT CHAIN 1140 1370 NS2A protein.  
 FT CHAIN 1371 1501 NS2B protein.  
 FT CHAIN 1502 2120 NS3 protein.  
 FT CHAIN 2121 2269 NS4A protein.  
 FT CHAIN 2270 2525 NS4B protein.  
 FT CHAIN 2526 3430 NS5 protein.  
 SQ SEQUENCE 3430 AA; 379866 MW; B03CBB31C86FD33B CRC64;

Query Match 97.5%; Score 2573; DB 2; Length 3430;  
 Best Local Similarity 98.2%; Pred. No. 2.4e-186;  
 Matches 492; Conservative 3; Mismatches 2; Indels 4; Gaps 1;

Qy	1	FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC	60
Db	291	FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC	350
Qy	61	YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA	120
Db	351	YLASVSDLSTRAACPTMGEAHNEKRADPAFVCKQGVVDRGWNGCGLFGKGSIDTCAKFA	410
Qy	121	CTTKATGWIIQKENIKYEVAI FVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL	180
Db	411	CTTKATGWIIQKENIKYEVAI FVHGPTTVESHG---KIGATQAGRFSITPSAPSYTLKL	466
Qy	181	GEYGEVTVDCEPRSGIDTSAYYVMSVGAKSFLVHREWFMMDLNLPWSSAGSTTWRNRETL	240
Db	467	GEYGEVTVDCEPRSGIDTSAYYVMSVGAKSFLVHREWFMMDLNLPWSSAGSTTWRNRETL	526
Qy	241	EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT	300
Db	527	EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT	586
Qy	301	TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPCKVPISSVASLNDLTPVGRLOTVNP	360
Db	587	TYGVCSKAFKFARTPADTGHGTVVLELQYTGDGPCKVPISSVASLNDLTPVGRLOTVNP	646
Qy	361	FVSVATANSKVLELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLRGAQRLAA	420
Db	647	FVSVATANSKVLELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLRGAQRLAA	706
Qy	421	LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARD	480
Db	707	LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARD	766
Qy	481	IAMTFLAVGGVLLFLSVNVHA	501
Db	767	IAMTFLAVGGVLLFLSVNVHA	787

RESULT 6  
POLG\_WNV  
ID POLG\_WNV STANDARD; PRT; 3430 AA.  
AC P06935;

DT 01-JAN-1988, integrated into UniProtKB/Swiss-Prot.  
DT 24-OCT-2003, sequence version 2.  
DT 07-MAR-2006, entry version 64.  
DE Genome polyprotein [Contains: Capsid protein C (Core protein);  
DE Envelope protein M (Matrix protein); Major envelope protein E;  
DE Nonstructural protein 1 (NS1); Nonstructural protein 2A (NS2A);  
DE Flavivirin protease NS2B regulatory subunit; Flavivirin protease NS3  
DE catalytic subunit (EC 3.4.21.91); Nonstructural protein 4A (NS4A);  
DE Nonstructural protein 4B (NS4B); RNA-directed RNA polymerase  
DE (EC 2.7.7.48) (NS5)].  
OS West Nile virus (WN).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Flavivirus; Japanese encephalitis virus group.  
OX NCBI\_TaxID=11082;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [GENOMIC RNA].  
RX MEDLINE=86124703; PubMed=3753811;  
RA Castle E., Leidner U., Nowak T., Wengler G., Wengler G.;  
RT "Primary structure of the West Nile flavivirus genome region coding  
RT for all nonstructural proteins.";  
RL Virology 149:10-26(1986).  
RN [2]  
RP SEQUENCE REVISION TO 1908; 2018-2036; 2242 AND 2859-2860.  
RX MEDLINE=21176376; PubMed=11277701; DOI=10.1006/viro.2000.0795;  
RA Yamshchikov V.F., Wengler G., Perelygin A.A., Brinton M.A.,  
RA Compans R.W.;  
RT "An infectious clone of the West Nile flavivirus.";  
RL Virology 281:294-304(2001).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [GENOMIC RNA] OF 1-291.  
RX MEDLINE=85274372; PubMed=2992152;  
RA Castle E., Nowak T., Leidner U., Wengler G., Wengler G.;  
RT "Sequence analysis of the viral core protein and the membrane-  
RT associated proteins V1 and NV2 of the flavivirus West Nile virus and  
RT of the genome sequence for these proteins.";  
RL Virology 145:227-236(1985).  
RN [4]  
RP NUCLEOTIDE SEQUENCE [GENOMIC RNA] OF 255-854.  
RX MEDLINE=86072082; PubMed=3855247;  
RA Wengler G., Castle E., Leidner U., Nowak T., Wengler G.;  
RT "Sequence analysis of the membrane protein V3 of the flavivirus West  
RT Nile virus and of its gene.";  
RL Virology 147:264-274(1985).  
RN [5]  
RP DISULFIDE BONDS IN E PROTEIN.  
RX MEDLINE=87122143; PubMed=3811228;  
RA Nowak T., Wengler G.;  
RT "Analysis of disulfides present in the membrane proteins of the West  
RT Nile flavivirus.";  
RL Virology 156:127-137(1987).  
CC -!- FUNCTION: The small proteins NS2A, NS4A and NS4B are hydrophobic,  
CC suggesting a possible membrane-related function. NS5 may play a  
CC role in the viral RNA replication. The NS2B/NS3 protease complex  
CC processes the viral polyprotein.  
CC -!- CATALYTIC ACTIVITY: Selective hydrolysis of -Xaa-Xaa-|-Yaa- bonds  
CC in which each of the Xaa can be either Arg or Lys and Yaa can be  
CC either Ser or Ala.

CC -!- CATALYTIC ACTIVITY: Nucleoside triphosphate + RNA(n) = diphosphate  
CC   + RNA(n+1).  
CC -!- SUBUNIT: NS3 and NS2B form a heterodimer. NS3 is the catalytic  
CC   subunit, whereas NS2B strongly stimulates the latter (By  
CC   similarity).  
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins  
CC   (By similarity).  
CC -!- MISCELLANEOUS: The virion of this virus is a nucleocapsid covered  
CC   by a lipoprotein envelope. The envelope contains two proteins: the  
CC   protein M and glycoprotein E. The nucleocapsid is a complex of  
CC   protein C and mRNA. In immature particles, there are 60  
CC   icosaedrally organized trimeric spikes on the surface. Each spike  
CC   consists of three heterodimers of envelope protein M precursor  
CC   (prM) and envelope protein E (By similarity).  
CC -!- SIMILARITY: Contains 1 peptidase S7 domain.  
CC -!- SIMILARITY: Contains 1 RdRp catalytic domain.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; M12294; AAA48498.2; -; Genomic\_RNA.  
DR PIR; A25256; GNWVWW.  
DR HSSP; Q88653; 1L9K.  
DR SMR; P06935; 25-97.  
DR MEROPS; S07.001; -.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH\_N.  
DR InterPro; IPR002464; DEAH\_box.  
DR InterPro; IPR011999; Flav\_glyE\_cen\_dm.  
DR InterPro; IPR001122; Flavi\_capsidC.  
DR InterPro; IPR011492; Flavi\_DEAD.  
DR InterPro; IPR000069; Flavi\_M.  
DR InterPro; IPR001157; Flavi\_NS1.  
DR InterPro; IPR000752; Flavi\_NS2A.  
DR InterPro; IPR000487; Flavi\_NS2B.  
DR InterPro; IPR000404; Flavi\_NS4A.  
DR InterPro; IPR001528; Flavi\_NS4B.  
DR InterPro; IPR000208; Flavi\_NS5.  
DR InterPro; IPR002535; Flavi\_propep.  
DR InterPro; IPR000336; Flv\_glyE\_Ig-like.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR001850; Peptidase\_S7.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR InterPro; IPR002877; RrmJFtsJ\_mtfrase.  
DR InterPro; IPR011998; Vrl\_glyE\_cen\_dim.  
DR Pfam; PF01003; Flavi\_capsid; 1.  
DR Pfam; PF07652; Flavi\_DEAD; 1.  
DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
DR Pfam; PF00869; Flavi\_glycoprot; 1.  
DR Pfam; PF01004; Flavi\_M; 1.  
DR Pfam; PF00948; Flavi\_NS1; 1.  
DR Pfam; PF01005; Flavi\_NS2A; 1.  
DR Pfam; PF01002; Flavi\_NS2B; 1.  
DR Pfam; PF01350; Flavi\_NS4A; 1.  
DR Pfam; PF01349; Flavi\_NS4B; 1.  
DR Pfam; PF00972; Flavi\_NS5; 1.

DR Pfam; PF01570; Flavi\_propep; 1.  
 DR Pfam; PF01728; FtsJ; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00949; Peptidase\_S7; 1.  
 DR ProDom; PD001496; Flavi\_NS1; 1.  
 DR SMART; SM00487; DEXDc; 1.  
 DR SMART; SM00490; HELICc; 1.  
 DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; FALSE\_NEG.  
 DR PROSITE; PS50507; RDRP\_SS RNA\_POS; 1.  
 KW ATP-binding; Capsid protein; Core protein; Envelope protein;  
 KW Glycoprotein; Helicase; Hydrolase; Membrane; Nucleotide-binding;  
 KW Nucleotidyltransferase; Polyprotein; RNA-directed RNA polymerase;  
 KW Structural protein; Transferase; Transmembrane.  
 FT CHAIN 1 123 Capsid protein C.  
 FT /FTId=PRO\_0000037743.  
 FT INIT\_MET 1 1 Removed from capsid protein C by the  
 FT cellular aminopeptidase.  
 FT PROPEP 124 215  
 FT /FTId=PRO\_0000037744.  
 FT CHAIN 216 290 Envelope protein M.  
 FT /FTId=PRO\_0000037745.  
 FT CHAIN 291 787 Major envelope protein E.  
 FT /FTId=PRO\_0000037746.  
 FT CHAIN 788 1139 Nonstructural protein 1.  
 FT /FTId=PRO\_0000037747.  
 FT CHAIN 1140 1370 Nonstructural protein 2A.  
 FT /FTId=PRO\_0000037748.  
 FT CHAIN 1371 1501 Flavivirin protease NS2B regulatory  
 FT subunit.  
 FT /FTId=PRO\_0000037749.  
 FT CHAIN 1502 2120 Flavivirin protease NS3 catalytic  
 FT subunit.  
 FT /FTId=PRO\_0000037750.  
 FT CHAIN 2121 2269 Nonstructural protein 4A.  
 FT /FTId=PRO\_0000037751.  
 FT CHAIN 2270 2525 Nonstructural protein 4B.  
 FT /FTId=PRO\_0000037752.  
 FT CHAIN 2526 3430 RNA-directed RNA polymerase.  
 FT /FTId=PRO\_0000037753.  
 FT DOMAIN 1508 1679 Peptidase S7.  
 FT DOMAIN 3055 3207 RdRp catalytic.  
 FT NP\_BIND 1695 1702 ATP (Potential).  
 FT REGION 388 401 Involved in fusion.  
 FT MOTIF 1786 1789 DEAH box.  
 FT ACT\_SITE 1552 1552 Charge relay system (By similarity).  
 FT ACT\_SITE 1576 1576 Charge relay system (By similarity).  
 FT ACT\_SITE 1636 1636 Charge relay system (By similarity).  
 FT CARBOHYD 138 138 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 917 917 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 962 962 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 994 994 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 1289 1289 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 2336 2336 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 2489 2489 N-linked (GlcNAc. . .) (Potential).  
 FT DISULFID 293 320  
 FT DISULFID 350 406  
 FT DISULFID 364 395

FT DISULFID 382 411  
 FT DISULFID 476 574  
 FT DISULFID 591 622  
 SQ SEQUENCE 3430 AA; 380110 MW; 42D71B7CB12DC45B CRC64;  
  
 Query Match 97.5%; Score 2572; DB 1; Length 3430;  
 Best Local Similarity 98.2%; Pred. No. 2.9e-186;  
 Matches 492; Conservative 2; Mismatches 3; Indels 4; Gaps 1;  
  
 Qy 1 FNCLGMSNRDFLEGVSGATWDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 60  
 |||||||  
 Db 291 FNCLGMSNRDFLEGVSGATWDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 350  
  
 Qy 61 YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWGNGCGLFGKGSIDTCAKFA 120  
 |||||||:|||||||:|||||||  
 Db 351 YLASVSDLSTRAACPTMGEAHNEKRADPAFVCKQGVVDRGWGNGCGLFGKGSIDTCAKFA 410  
  
 Qy 121 CTTKATGWIIQKENIKYEVAIFVHGPTTVEHGNYSTQIGATQAGRFSITPSAPSYTLKL 180  
 |||||||:||||||| :|||||||  
 Db 411 CTTKATGWIIQKENIKYEVAIFVHGPTTVEHG---KIGATQAGRFSITPSAPSYTLKL 466  
  
 Qy 181 GEYGEVTVDCEPRSGIDTSAYYVMSVGAFLVHREWMDLNLPWSSAGSTWRNRETL 240  
 |||||||  
 Db 467 GEYGEVTVDCEPRSGIDTSAYYVMSVGEKFLVHREWMDLNLPWSSAGSTWRNRETL 526  
  
 Qy 241 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 300  
 |||||||  
 Db 527 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 586  
  
 Qy 301 TYGVCSKAFKFARTPADTGHTVVLELQYTGKDGPCKVPISSVASLNDLTPVGRLVTVNP 360  
 |||||||  
 Db 587 TYGVCSKAFKFARTPADTGHTVVLELQYTGDGPCKVPISSVASLNDLTPVGRLVTVNP 646  
  
 Qy 361 FVSVATANSKVLELEPPFSDSYIVVGRGEQQINHHWHKGSSIGKAFTTLRGAQRLAA 420  
 |||||||  
 Db 647 FVSVATANSKVLELEPPFGDSYIVVGRGEQQINHHWHKGSSIGKAFTTLRGAQRLAA 706  
  
 Qy 421 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARD 480  
 |||||||  
 Db 707 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARD 766  
  
 Qy 481 IAMTFLAVGGVLLFLSVNVHA 501  
 |||||||  
 Db 767 IAMTFLAVGGVLLFLSVNVHA 787

RESULT 7  
 Q5EVN3\_WNV  
 ID Q5EVN3\_WNV PRELIMINARY; PRT; 3433 AA.  
 AC Q5EVN3;  
 DT 15-MAR-2005, integrated into UniProtKB/TrEMBL.  
 DT 15-MAR-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 5.  
 DE Polyprotein.  
 OS West Nile virus (WN).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus; Japanese encephalitis virus group.

OX NCBI\_TaxID=11082;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=96-111;  
RX PubMed=15752452;  
RA Schuffenecker I., Peyrefitte C.N., el Harrak M., Murri S., Leblond A.,  
RA Zeller H.G.;  
RT "West Nile Virus in Morocco, 2003.";  
RL Emerg. Infect. Dis. 11:306-309(2005).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AY701412; AAT92098.1; -; Genomic\_RNA.  
DR SMR; Q5EVN3; 25-97.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003725; F:double-stranded RNA binding; IEA.  
DR GO; GO:0003724; F:RNA helicase activity; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH\_N.  
DR InterPro; IPR011999; Flav\_glyE\_cen\_dm.  
DR InterPro; IPR001122; Flavi\_capsidC.  
DR InterPro; IPR011492; Flavi\_DEAD.  
DR InterPro; IPR000069; Flavi\_M.  
DR InterPro; IPR001157; Flavi\_NS1.  
DR InterPro; IPR000752; Flavi\_NS2A.  
DR InterPro; IPR000487; Flavi\_NS2B.  
DR InterPro; IPR000404; Flavi\_NS4A.  
DR InterPro; IPR001528; Flavi\_NS4B.  
DR InterPro; IPR000208; Flavi\_NS5.  
DR InterPro; IPR002535; Flavi\_propep.  
DR InterPro; IPR000336; Flv\_glyE\_Ig-like.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR001850; Peptidase\_S7.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR InterPro; IPR002877; RrmJFtsJ\_mtfrase.  
DR InterPro; IPR011998; Vrl\_glyE\_cen\_dim.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF01003; Flavi\_capsid; 1.  
DR Pfam; PF07652; Flavi\_DEAD; 1.  
DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
DR Pfam; PF00869; Flavi\_glycoprot; 1.  
DR Pfam; PF01004; Flavi\_M; 1.  
DR Pfam; PF00948; Flavi\_NS1; 1.  
DR Pfam; PF01005; Flavi\_NS2A; 1.  
DR Pfam; PF01002; Flavi\_NS2B; 1.  
DR Pfam; PF01350; Flavi\_NS4A; 1.  
DR Pfam; PF01349; Flavi\_NS4B; 1.

DR Pfam; PF00972; Flavi\_NS5; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 DR Pfam; PF01728; FtsJ; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00949; Peptidase\_S7; 1.  
 DR ProDom; PD001496; Flavi\_NS1; 1.  
 DR SMART; SM00487; DEXDc; 1.  
 DR SMART; SM00490; HELICc; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1..  
 KW Polyprotein.  
 SQ SEQUENCE 3433 AA; 381249 MW; 7ECC96DBFD9D53DA CRC64;

Query Match 96.0%; Score 2532; DB 2; Length 3433;  
 Best Local Similarity 95.6%; Pred. No. 3.2e-183;  
 Matches 479; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Qy	1 FNCLGMSNRDFLEGVSGATWVDLVLEGSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC	60
	:	
Db	291 FNCLGMSNRDFLEGVSGATWVDLVLEGSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYC	350
Qy	61 YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWGNGCGLFGKGSIDTCAKFA	120
	:         :     :     :	
Db	351 YLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWGNGCGLFGKGSIDTCAKFA	410
Qy	121 CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL	180
	:               :     :     :     :	
Db	411 CSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPAAPSYTLKL	470
Qy	181 GEYGEVTVDCEPRSGIDTSAYYVMSVGAFLVHREWFMDLNLPWSSAGSTWRNRETLM	240
	:     :    :     :     :	
Db	471 GEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWFMDLNLPWSSAGSTVWRNRETLM	530
Qy	241 EFEPEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT	300
	:     :     :     :     :	
Db	531 EFEPEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT	590
Qy	301 TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPCKVPISSVASLNDLTPVGRLOTVNP	360
	:     :     :	
Db	591 TYGVCSKAFKFLGTPADTGHGTVVLELQYTGDGPCKVPISSVASLNDLTPVGRLOTVNP	650
Qy	361 FVSVATANSKVLELEPPFSDSYIVVGRGEQQINHHWHKGSSIGKAFTTLRGAQRLAA	420
	:     :     :     :     :	
Db	651 FVSVATANAKVLELEPPFGDSYIVVGRGEQQINHHWHKGSSIGKAFTTLKGAAQRLAA	710
Qy	421 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS	480
	:     :     :     :     :	
Db	711 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS	770
Qy	481 IAMTFLAVGGVLLFLSVNVHA 501	
	:     :	
Db	771 IALTFLAVGGVLLFLSVNVHA 791	

RESULT 8  
 Q9WI84\_WNV  
 ID Q9WI84\_WNV PRELIMINARY; PRT; 501 AA.  
 AC Q9WI84;

DT 01-NOV-1999, integrated into UniProtKB/TrEMBL.  
DT 01-NOV-1999, sequence version 1.  
DT 07-FEB-2006, entry version 20.  
DE Envelope glycoprotein (Fragment).  
OS West Nile virus (WN).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Flavivirus; Japanese encephalitis virus group.  
OX NCBI\_TaxID=11082;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=KN3829;  
RX MEDLINE=20271587; PubMed=10813479;  
RA Miller B.R., Naschi R.S., Godsey M.S., Savage H.M., Lutwama J.J.,  
RA Lanciotti R.S., Peters C.J.;  
RT "First field evidence for natural vertical transmission of West Nile  
RT virus in Culex univittatus complex mosquitoes from Rift Valley  
RT province, Kenya.";  
RL Am. J. Trop. Med. Hyg. 62:240-246(2000).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AF146082; AAD31720.1; -; Genomic\_RNA.  
DR HSSP; Q88653; 1OKE.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR011999; Flav\_glyE\_cen\_dm.  
DR InterPro; IPR000336; Flv\_glyE\_Ig-like.  
DR InterPro; IPR011998; Vrl\_glyE\_cen\_dim.  
DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
DR Pfam; PF00869; Flavi\_glycoprot; 1.  
KW Envelope protein.  
FT NON\_TER 1 1  
FT NON\_TER 501 501  
SQ SEQUENCE 501 AA; 53622 MW; D2A9C827F71C00D5 CRC64;  
  
Query Match 95.9%; Score 2531; DB 2; Length 501;  
Best Local Similarity 95.4%; Pred. No. 2.8e-184;  
Matches 478; Conservative 14; Mismatches 9; Indels 0; Gaps 0;  
  
Qy 1 FNCLGMSNRDFLEGVSGATWDLVLEGDSCTIMSKDKPTIDVKMMNMEAANLADVRSYC 60  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 1 FNCLGMSNRDFLEGVSGATWDLVLEGDSCTIMSKDKPTIDVKMMNMEAANLAEVRSYC 60  
  
Qy 61 YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWGNCGGLFGKGSIDTCAKFA 120  
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 61 YLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWGNCGGLFGKGSIDTCAKFA 120  
  
Qy 121 CTTKATGWIIQKENIKYEVAIFVHGPTTVEHGNYSTQIGATQAGRFSITPSAPSYTLKL 180  
|:||||| | |||||||:|||||:|||||:|||||:  
Db 121 CSTKATGRTILKENIKYEVAIFVHGPTTVEHGNYSTQIGATQAGRFSITPAAPSYTLKL 180  
  
Qy 181 GEYGEVTVDCEPRSGIDTSAYYVMSVGAKSFLVHREWFDLNLPWSSAGSTWRNRETL 240  
|||:|||||:|||||:|||||:  
Db 181 GEYGEVTVDCEPRSGIDTNAYYVMTVGTFLVHREWFDLNLPWSSAGSTVWRNRETL 240

Qy 241 EFEEPHATKQSVALGSQEGALHQALAGAIPVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 300  
 |||||||:|||||||:  
 Db 241 EFEEPHATKQSVALGSQEGALHQALAGAIPVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 300  
 |||||||:  
 Qy 301 TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPCKVPISSVASLNDLTPVGRLVNP 360  
 |||||||:  
 Db 301 TYGVCSKAFKFGLGTPADTGHGTVVLELQYTGDGPCKVPISSVASLNDLTPVGRLVNP 360  
 |||||||:  
 Qy 361 FVSVATANSKVIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLRGAQRLAA 420  
 |||||||:  
 Db 361 FVSVATANAKVIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLKGAQRLAA 420  
 |||||||:  
 Qy 421 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS 480  
 |||||||:  
 Db 421 LGDTAWDFGSVGGVFTSVGKAVHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS 480  
 |||||||:  
 Qy 481 IAMTFLAVGGVLLFLSVNVHA 501  
 ||:|||||:  
 Db 481 IALTFLAVGGVLLFLSVNVHA 501

#### RESULT 9

**Q9WHD1\_WNV**  
 ID Q9WHD1\_WNV PRELIMINARY; PRT; 773 AA.  
 AC Q9WHD1;  
 DT 01-NOV-1999, integrated into UniProtKB/TrEMBL.  
 DT 01-NOV-1999, sequence version 1.  
 DT 07-FEB-2006, entry version 24.  
 DE Polyprotein (Fragment).  
 OS West Nile virus (WN).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus; Japanese encephalitis virus group.  
 OX NCBI\_TaxID=11082;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=RO97-50;  
 RX MEDLINE=20014331; PubMed=10548295;  
 RA Savage H.M., Ceianu C., Nicolescu G., Karabatsos N., Lanciotti R.,  
 RA Vladimirescu A., Laiv L., Ungureanu A., Romanca C., Tsai T.F.;  
 RT "Entomologic and avian investigations of an epidemic of West Nile  
 RT fever in Romania in 1996, with serologic and molecular  
 RT characterization of a virus isolate from mosquitoes.";  
 RL Am. J. Trop. Med. Hyg. 61:600-611(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=RO97-50;  
 RA Lanciotti R.L., Ludwig M.L., Savage H.M.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; AF130362; AAD28623.1; -; Genomic\_RNA.  
 DR HSSP; Q88653; 1OKE.  
 DR SMR; Q9WHD1; 1-72.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0019058; P:viral infectious cycle; IEA.  
 DR InterPro; IPR011999; Flav\_glyE\_cen\_dm.  
 DR InterPro; IPR001122; Flavi\_capsidC.  
 DR InterPro; IPR000069; Flavi\_M.  
 DR InterPro; IPR002535; Flavi\_propep.  
 DR InterPro; IPR000336; Flv\_glyE\_Ig-like.  
 DR InterPro; IPR011998; Vrl\_glyE\_cen\_dim.  
 DR Pfam; PF01003; Flavi\_capsid; 1.  
 DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
 DR Pfam; PF00869; Flavi\_glycoprot; 1.  
 DR Pfam; PF01004; Flavi\_M; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 KW Polyprotein.  
 FT CHAIN <1 88 capsid protein.  
 FT CHAIN 89 265 pre-membrane/membrane protein.  
 FT CHAIN 266 766 envelope glycoprotein.  
 FT NON\_TER 1 1  
 FT NON\_TER 773 773  
 SQ SEQUENCE 773 AA; 83364 MW; 2C33EA27EC676EE7 CRC64;  
  
 Query Match 95.9%; Score 2531; DB 2; Length 773;  
 Best Local Similarity 95.4%; Pred. No. 5.1e-184;  
 Matches 478; Conservative 14; Mismatches 9; Indels 0; Gaps 0;  
  
 Qy 1 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 60  
 |||||||  
 Db 266 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYC 325  
  
 Qy 61 YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWGNCGLFGKGSIDTCAKFA 120  
 |||:  
 Db 326 YLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWGNCGLFGKGSIDTCAKFA 385  
  
 Qy 121 CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL 180  
 :|||||| ||||||  
 Db 386 CSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPAAPSYTLKL 445  
  
 Qy 181 GEYGEVTVDCEPRSGIDTSAYYMSVGAKSFLVHREWMDLNLPWSSAGSTWRNRETL 240  
 |||||||:|||||:||| :|||||  
 Db 446 GEYGEVTVDCEPRSGIDTNAYYMTVGTFLVHREWMDLNLPWSSAGSTVWRNRETL 505  
  
 Qy 241 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 300  
 |||||||:|||||  
 Db 506 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 565  
  
 Qy 301 TYGVCSKAFKFARTPADTGHGTVVLEQYTGKDGPCKVPISSVASLNDLTPVGRLVTVNP 360  
 |||||||  
 Db 566 TYGVCSKAFKFLGT PADTGHGTVVLEQYTGDGPCKVPISSVASLNDLTPVGRLVTVNP 625  
  
 Qy 361 FVSVATANSKVLELEPPFSDSYIVVGRGEQQINHHWHKGSSSIGKAFTTTLRGAQRLAA 420  
 |||||||:  
 Db 626 FVSVATANAKVLELEPPFGDSYIVVGRGEQQINHHWHKGSSSIGKAFTTTLKGAQRLAA 685  
  
 Qy 421 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS 480  
 |||||||:  
 Db 686 LGDTAWDFGSVGGVFTSVGKAVHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS 745

QY 481 IAMTFLAVGGVLLFLSVNVHA 501  
| : | : | : | : | : | : | : |  
Db 746 IALTFLAVGGVLLFLSVNVHA 766

RESULT 10

Q5EVN2\_WNV

ID Q5EVN2\_WNV PRELIMINARY; PRT; 3433 AA.  
AC Q5EVN2;  
DT 15-MAR-2005, integrated into UniProtKB/TrEMBL.  
DT 15-MAR-2005, sequence version 1.  
DT 07-FEB-2006, entry version 5.  
DE Polyprotein.  
OS West Nile virus (WN).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Flavivirus; Japanese encephalitis virus group.  
OX NCBI\_TaxID=11082;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=04.05;  
RX PubMed=15752452;  
RA Schuffenecker I., Peyrefitte C.N., el Harrak M., Murri S., Leblond A.,  
RA Zeller H.G.;  
RT "West Nile Virus in Morocco, 2003.";  
RL Emerg. Infect. Dis. 11:306-309(2005).

CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----

DR EMBL; AY701413; AAT92099.1; -; Genomic\_RNA.  
DR SMR; Q5EVN2; 25-97.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003725; F:double-stranded RNA binding; IEA.  
DR GO; GO:0003724; F:RNA helicase activity; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH\_N.  
DR InterPro; IPR011999; Flav\_glyE\_cen\_dm.  
DR InterPro; IPR001122; Flavi\_capsidC.  
DR InterPro; IPR011492; Flavi\_DEAD.  
DR InterPro; IPR000069; Flavi\_M.  
DR InterPro; IPR001157; Flavi\_NS1.  
DR InterPro; IPR000752; Flavi\_NS2A.  
DR InterPro; IPR000487; Flavi\_NS2B.  
DR InterPro; IPR000404; Flavi\_NS4A.  
DR InterPro; IPR001528; Flavi\_NS4B.  
DR InterPro; IPR000208; Flavi\_NS5.  
DR InterPro; IPR002535; Flavi\_propep.  
DR InterPro; IPR000336; Flv\_glyE\_Ig-like.

DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR001850; Peptidase\_S7.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR InterPro; IPR002877; RrmJFtsJ\_mtfrase.  
 DR InterPro; IPR011998; Vrl\_glyE\_cen\_dim.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF01003; Flavi\_capsid; 1.  
 DR Pfam; PF07652; Flavi\_DEAD; 1.  
 DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
 DR Pfam; PF00869; Flavi\_glycoprot; 1.  
 DR Pfam; PF01004; Flavi\_M; 1.  
 DR Pfam; PF00948; Flavi\_NS1; 1.  
 DR Pfam; PF01005; Flavi\_NS2A; 1.  
 DR Pfam; PF01002; Flavi\_NS2B; 1.  
 DR Pfam; PF01350; Flavi\_NS4A; 1.  
 DR Pfam; PF01349; Flavi\_NS4B; 1.  
 DR Pfam; PF00972; Flavi\_NS5; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 DR Pfam; PF01728; FtsJ; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00949; Peptidase\_S7; 1.  
 DR ProDom; PD001496; Flavi\_NS1; 1.  
 DR SMART; SM00487; DEXDc; 1.  
 DR SMART; SM00490; HELICc; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 KW Polyprotein.  
 SQ SEQUENCE 3433 AA; 381202 MW; A98222C50069232A CRC64;  
  
 Query Match 95.9%; Score 2531; DB 2; Length 3433;  
 Best Local Similarity 95.4%; Pred. No. 3.8e-183;  
 Matches 478; Conservative 14; Mismatches 9; Indels 0; Gaps 0;  
  
 Qy 1 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 60  
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 Db 291 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYC 350  
  
 Qy 61 YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWGNCGLFGKGSIDTCAKFA 120  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:  
 Db 351 YLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWGNCGLFGKGSIDTCAKFA 410  
  
 Qy 121 CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL 180  
 |:|||||:|||:|||||:|||||:|||||:|||||:  
 Db 411 CSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPAAPSYTLKL 470  
  
 Qy 181 GEYGEVTVDCEPRSGIDTSAYYMSVGAKSFLVHREWMDLNLPWSSAGSTTWRNRETL 240  
 |||||||:|||||:|||||:|||:|||||:  
 Db 471 GEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWMDLNLPWSSAGSTVWRNRETL 530  
  
 Qy 241 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 300  
 |||||||:|||||:  
 Db 531 EFEEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVKMEKLQLKGT 590  
  
 Qy 301 TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPCKVPISSVASLNDLTPVGRLLTVNP 360  
 |||||||:  
 Db 591 TYGVCSKAFKFLGTPADTGHGTVVLELQYTGDGPCKVPISSVASLNDLTPVGRLLTVNP 650

Qy 361 FVSVATANSKVLIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLRGAQRLAA 420  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 651 FVSVATANAKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLKGAQRLAA 710  
  
Qy 421 LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS 480  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 711 LGDTAWDFGSVGGVFTSVGKAVHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS 770  
  
Qy 481 IAMTFLAVGGVLLFLSVNVHA 501  
||:|||||:|||||:  
Db 771 IALTFLAVGGVLLFLSVNVHA 791

RESULT 11

Q6WV07\_WNV

ID Q6WV07\_WNV PRELIMINARY; PRT; 3433 AA.  
AC Q6WV07;  
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2004, sequence version 1.  
DT 07-FEB-2006, entry version 8.  
DE Polyprotein.  
GN Name=pol;  
OS West Nile virus (WN).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Flavivirus; Japanese encephalitis virus group.  
OX NCBI\_TaxID=11082;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=PaAn001;  
RX MEDLINE=22949215; PubMed=14585341; DOI=10.1016/S0042-6822(03)00536-1;  
RA Charrel R.N., Brault A.C., Gallian P., Lemasson J.-J., Murgue B.,  
RA Murri S., Pastorino B., Zeller H., de chesse R., de Micco P.,  
RA de Lamballerie X.;  
RT "Evolutionary relationship between Old World West Nile virus strains.  
RT Evidence for viral gene flow between Africa, the Middle East, and  
RT Europe.";  
RL Virology 315:381-388 (2003).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AY268132; AAQ00998.1; -; Genomic\_RNA.  
DR HSSP; Q9Q4T1; 1BEE.  
DR SMR; Q6WV07; 25-97.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003725; F:double-stranded RNA binding; IEA.  
DR GO; GO:0003724; F:RNA helicase activity; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH\_N.

DR InterPro; IPR011999; Flav\_glyE\_cen\_dm.  
 DR InterPro; IPR001122; Flavi\_capsidC.  
 DR InterPro; IPR011492; Flavi\_DEAD.  
 DR InterPro; IPR000069; Flavi\_M.  
 DR InterPro; IPR001157; Flavi\_NS1.  
 DR InterPro; IPR000752; Flavi\_NS2A.  
 DR InterPro; IPR000487; Flavi\_NS2B.  
 DR InterPro; IPR000404; Flavi\_NS4A.  
 DR InterPro; IPR001528; Flavi\_NS4B.  
 DR InterPro; IPR000208; Flavi\_NS5.  
 DR InterPro; IPR002535; Flavi\_propep.  
 DR InterPro; IPR000336; Flv\_glyE\_Ig-like.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR001850; Peptidase\_S7.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR InterPro; IPR002877; RrmJFtsJ\_mtfrase.  
 DR InterPro; IPR011998; Vrl\_glyE\_cen\_dim.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF01003; Flavi\_capsid; 1.  
 DR Pfam; PF07652; Flavi\_DEAD; 1.  
 DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
 DR Pfam; PF00869; Flavi\_glycoprot; 1.  
 DR Pfam; PF01004; Flavi\_M; 1.  
 DR Pfam; PF00948; Flavi\_NS1; 1.  
 DR Pfam; PF01005; Flavi\_NS2A; 1.  
 DR Pfam; PF01002; Flavi\_NS2B; 1.  
 DR Pfam; PF01350; Flavi\_NS4A; 1.  
 DR Pfam; PF01349; Flavi\_NS4B; 1.  
 DR Pfam; PF00972; Flavi\_NS5; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 DR Pfam; PF01728; FtsJ; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00949; Peptidase\_S7; 1.  
 DR ProDom; PD001496; Flavi\_NS1; 1.  
 DR SMART; SM00487; DEXDc; 1.  
 DR SMART; SM00490; HELICc; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 KW Polyprotein.  
 SQ SEQUENCE 3433 AA; 381104 MW; 2F25A8012B297680 CRC64;  
  
 Query Match 95.9%; Score 2531; DB 2; Length 3433;  
 Best Local Similarity 95.4%; Pred. No. 3.8e-183;  
 Matches 478; Conservative 14; Mismatches 9; Indels 0; Gaps 0;  
  
 Qy 1 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLADVRSYC 60  
 |||||||  
 Db 291 FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYC 350  
  
 Qy 61 YLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWGNCGLFGKGSIDTCAKFA 120  
 ||:|||||  
 Db 351 YLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWGNCGLFGKGSIDTCAKFA 410  
  
 Qy 121 CTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPSAPSYTLKL 180  
 |:|||||  
 Db 411 CSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPAAPSYTLKL 470

Qy	181	GEYGEVTVDCEPRSGIDTSAYVMSVGAKSFLVHREWFMIDLNLPWSSAGSTTWRNRETLM	240
		:     :    :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	471	GEYGEVTVDCEPRSGIDTNAYVMTVGTKTFLVHREWFMIDLNLPWSSAGSTVWRNRETLM	530
Qy	241	EFEPEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVMEKLQLKGT	300
		:     :     :     :     :     :     :     :     :     :     :	
Db	531	EFEPEPHATKQSVALGSQEGALHQALAGAI PVEFSSNTVKLTSGHLKCRVMEKLQLKGT	590
Qy	301	TYGVCSKAFKFARTPADTGHGTVVLELQYTGKDGPCKVPISSVASLNDLTPVGRRLVTVNP	360
		:     :     :     :     :     :     :     :     :     :     :	
Db	591	TYGVCSKAFKFGLTPADTGHGTVVLELQYTGDGPCKVPISSVASLNDLTPVGRRLVTVNP	650
Qy	361	FVSVATANSKVIELEPPFSDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLRGAQRLAA	420
	:	:     :     :     :     :     :     :     :     :     :	
Db	651	FVSVATANAKVIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLKGAQRLAA	710
Qy	421	LGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS	480
	:	:     :     :     :     :     :     :     :     :     :	
Db	711	LGDTAWDFGSVGGVFTSVGKAVHQVFGGAFRSLFGGMSWITQGLLGALLWMGINARDRS	770
Qy	481	IAMTFLAVGGVLLFLSVNVHA	501
	:	:	
Db	771	IALTFLAVGGVLLFLSVNVHA	791

#### RESULT 12

Q80B10\_WNV

ID Q80B10\_WNV PRELIMINARY; PRT; 3433 AA.  
AC Q80B10;  
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.  
DT 01-JUN-2003, sequence version 1.  
DT 07-FEB-2006, entry version 13.  
DE Polyprotein.  
OS West Nile virus (WN).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Flavivirus; Japanese encephalitis virus group.  
OX NCBI\_TaxID=11082;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=KN3829;  
RX MEDLINE=22949215; PubMed=14585341; DOI=10.1016/S0042-6822(03)00536-1;  
RA Charrel R.N., Brault A.C., Gallian P., Lemasson J.-J., Murgue B.,  
RA Murri S., Pastorino B., Zeller H., de chesse R., de Micco P.,  
RA de Lamballerie X.;  
RT "Evolutionary relationship between Old World West Nile virus strains.  
RT Evidence for viral gene flow between Africa, the Middle East, and  
RT Europe.";  
RL Virology 315:381-388(2003).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AY262283; AAP20887.1; -; Genomic\_RNA.  
DR HSSP; Q88653; 1L9K.  
DR SMR; Q80B10; 25-97.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.